Author:
Uren, Alan

Title:
The development of a novel patient reported outcome measure for the assessment of the signs, symptoms and impact of underactive bladder

The development of a novel patient reported outcome measure for the assessment of the signs, symptoms and impact of underactive bladder

Alan Uren

A thesis submitted to the University of Bristol in accordance with the requirements of the degree of PhD in the Faculty of the Bristol Medical School

May 2018

Word count: 48,798

Abstract

Underactive bladder (UAB) is a health issue which is receiving increasing attention in the urological literature. The most recent symptomatic definition approved by the International Continence Society steering committee in 2016 states: “Underactive bladder is characterised by a slow urinary stream, hesitancy and straining to void, with or without a feeling of incomplete bladder emptying and dribbling, often with storage symptoms”. UAB is considered a symptom syndrome suggestive of the urodynamic observation of detrusor underactivity (DU). The symptomatic burden of lower urinary tract symptoms (LUTS) associated with DU and known impact of LUTS on quality of life highlight the requirement to understand how the patient with UAB feels and functions for clinical outcome assessment purposes. Currently, no fully validated patient reported outcome (PRO) measures exist for the assessment of UAB. The thesis describes the development of the ICIQ-UAB, a new PRO measure for the assessment of the symptoms of UAB, and their associated bother and impact. Qualitative methodology was employed to understand how the clinical diagnosis of DU manifests as symptoms, by a thorough exploration of the lived experience of patients. Decisions on the inclusion of draft questionnaire items, including their content, language and response items, were made on the basis of the qualitative evidence and consultation with an expert clinical panel. Draft items were refined by cognitive interviews which confirmed the items to be understood and interpreted as intended by patients. Validity and test-retest reliability of the ICIQ-UAB was supported by a European pilot study, and the wider cultural applicability by additional patient interviews in the US and Japan. The developmental ICIQ-UAB is now ready for further large-scale validation in future clinical trials and is envisaged as an important tool for the monitoring of future UAB treatment strategies.

Acknowledgements

Firstly, I would like to thank Professor Paul Abrams, Professor Marcus Drake and Dr Nikki Cotterill for the opportunity to undertake this PhD. In particular, Dr Nikki Cotterill and Professor Marcus Drake have provided continuous and resolute supervisory support and expert guidance throughout. Their willingness and enthusiasm for discussions and direction when required, and the investment of time and effort in providing constructive feedback on drafts is hugely appreciated. I also extend my thanks to all the staff at the Bristol Urological Institute. It is a pleasure to work in such a friendly and supportive environment.

There have been a large number of people involved in all aspects of this project, including health and research professionals from multiple sites and hospitals, to whom I am very grateful. I would particularly like to thank the team at Astellas; Monique Klaver, Zalmai Hakimi, Dominique Bongaerts and Ton Kos, for their expertise and management of the outsourced aspects of the project. My gratitude also goes to the participants involved in all the sub-studies and interviews. It was pleasure to meet and to speak with them.

Finally, the support of my close family and fiancée (soon to be wife!) is continuous and for which I am forever thankful. The completion of this thesis has been an invaluable experience, and one which I will be proud of for life.

Author's Declaration

I declare that the work in this dissertation was carried out in accordance with the requirements of the University's Regulations and Code of Practice for Taught Programmes and that it has not been submitted for any other academic award. Except where indicated by specific reference in the text, this work is my own work. Work done in collaboration with, or with the assistance of others, is indicated as such. I have identified all material in this dissertation which is not my own work through appropriate referencing and acknowledgement. Where I have quoted or otherwise incorporated material which is the work of others, I have included the source in the references. Any views expressed in the dissertation, other than referenced material, are those of the author.

SIGNED: ……. DATE: ……..

1. Table of Contents

List of Tables

List of Figures

List of Publications

A. Uren and M. Drake (2017) Definition and Symptoms of Underactive Bladder. Invest Clin Urol. 58.

A. Uren, N. Cotterill, C. Harding, C. Hillary, C. Chapple, M. Klaver, D. Bongaerts, Z. Hakimi, P. Abrams (2017). Qualitative exploration of the patient experience of underactive bladder. Eur Urol. 72(3):402-407.

A. Uren, N. Cotterill, C. Harding, C. Hillary, C. Chapple, M. Klaver, D. Bongaerts, Z. Hakimi, P. Abrams (2017). Reply from Authors re: Mikkel Fode, Jens Sønksen. Towards a Greater Understanding of Underactive Bladder. Eur Urol. 72(3): 409-410.

A. Uren, N. Cotterill, C. Harding, C. Hillary, C. Chapple, M. Klaver, D. Bongaerts, Z. Hakimi, P. Abrams (2018). Reply to Bora Lee and Jae Heon Kim's Letter to the Editor re: Alan D. Uren, Nikki Cotterill, Christopher Harding, et al. Qualitative Exploration of the Patient Experience of Underactive Bladder. Eur Urol. 73(1) e15-e16.

The manuscript entitled 'The development of the ICIQ-UAB: a patient reported outcome measure for underactive bladder' was submitted to Neurourology and Urodynamics in May 2018, and is accepted pending minor revisions.

Funding and outsourcing

Astellas Pharma Europe funded this work and was involved in the design and management of the study, including the outsourcing of the qualitative work in the United States and Japan as detailed in the thesis. The author received a salary from the Bristol Urological Institute as a research assistant, who also funded the PhD tuition fees.

Chapter 1 The pathophysiology and epidemiology of underactive bladder

1.1 Introduction

The statement "the bladder is an unreliable witness" was first made by Bates in 1970, and refers to the complex problems urologists have when trying to interpret the symptoms presented by patients [1]. This introductory chapter presents a symptom syndrome that is increasingly recognised as being associated with bothersome symptoms of the lower urinary tract; the underactive bladder. The current status of the terminology surrounding underactive bladder is examined, and the known epidemiology, pathophysiology, diagnosis and treatment are discussed. Some content of this chapter was published in the Investigative and Clinical Urology as an invited review article entitled 'Definition and Symptoms of Underactive Bladder' in September, 2017 [2].

1.2 Lower urinary tract symptoms

The term lower urinary tract symptoms (LUTS) was introduced in 1994 [3], and relates to the symptoms experienced by individuals with pathology of the bladder, prostate (in men), and the urethra. LUTS are broadly divided in to three groups: storage, voiding and postmicturition symptoms [4]. Storage symptoms are experienced when there is a dysfunction in the ability of the bladder to store urine, such as increased urinary frequency, urgency, urinary incontinence, and nocturia. Voiding symptoms relate to problems with the flow of urine and include symptoms such as a slow/reduced stream, hesitancy and intermittency. Postmicturition symptoms encompass problems encountered immediately after voiding, such as the sensation of incomplete emptying or postmicturition dribble. There is a high prevalence of LUTS in the community. In a large population-based survey, the prevalence of at least one storage LUTS was found to be 51.3% in men and 59.2% in women. The prevalence of at least one voiding LUTS was less (16.9% in men and 14.2% in women) [5]. The prevalence of LUTS also increases with age [6], with 80% of men having benign prostate hyperplasia (BPH) and associated LUTS when over 80 years of age [7]. The known impact of LUTS on day to day life, such as the associated embarrassment or inconvenience of incontinence and increased urinary frequency on social, physical activities, and lifestyle, highlights the importance of meaningful treatment outcomes for these patients [8,9].

1.3 Urodynamics

A reliable method of evaluating bladder function is by urodynamic study [10]. The simplest of urodynamic techniques is non-invasive uroflowmetry, where urine flow is studied using relatively simple equipment to measure parameters such as the urinary flow rate per unit time expressed in millilitres per second (ml/s), maximum flow rate Q_{max} (ml/s), total voided volume VV (ml) and average flow rate Q_{ave} (ml/s). Abnormal flow rates, and flow patterns such as an intermittent stream can be interpreted and classified by different urological diagnoses. Further investigation is by cystometry, which is the method by which both the voiding and storage phases of micturition may be evaluated. During standard cystometry, the (intravesical) pressure p_{ves} in the bladder is measured along with the pressure in the abdominal cavity p_{abd} (usually in the rectum) during bladder filling and voiding [10]. The detrusor pressure p_{det} can then be calculated electronically ($p_{det}= p_{ves}-p_{abd}$) and other parameters such as the detrusor pressure at maximum flow $p_{det}Q_{max}$,. When the results of these investigations are related to the symptomatic complaints of the patient, urodynamic diagnoses can be made and treatment strategies planned accordingly.

1.4 Symptom syndromes for LUTS

There are common linkages of LUTS which can be used to derive symptom syndromes. The most widely recognised of these is the overactive bladder (OAB), defined as “urgency, with or without urge urinary incontinence, usually with increased daytime frequency and nocturia” [4,11]. The use of a symptom syndrome enables wider clinical recognition of the presenting symptoms, and identifies individuals who will potentially incur benefit from a specific intervention. Terminology such as ‘overactive bladder’ also facilitates patient understanding of their symptoms, as their understanding intuitively fits with how they perceive their own symptoms. When investigated further by healthcare professionals, patients with OAB are sometimes identified as having detrusor overactivity (DO). This is a urodynamic observation of bladder contraction during filling, which may be spontaneous or provoked [12]. Alternatively, there may be findings of inflammation in the lower urinary tract, or some other factor which sensitises the sensory nerves from the lower urinary tract. Thus, storage LUTS are largely encompassed by a symptom syndrome (OAB) and urodynamic observation (DO). These two terms are not interchangeable, since OAB patients may not have DO on filling cystometry; likewise DO may occur with no associated urgency [13]. A similar approach based on a symptom syndrome and urodynamic observation, has been proposed in further developing the field for voiding LUTS. Detrusor underactivity (DU), an increasingly recognised cause of troublesome

voiding, is considered the analogous urodynamic observation to the proposed underactive bladder (UAB) symptom complex [14,15]. Although overactive bladder syndrome is comparatively well understood, underactive bladder syndrome is relatively under-researched.

1.5 Terminology of underactive bladder

The International Continence Society (ICS) derived the terminology of lower urinary tract function [4] and is the main basis for the wordings and definitions used most widely in clinical practice. There has been much interest generated in this topic in recent years and efforts made to reach professional consensus on symptomatic definitions.

The definition of urodynamically diagnosed DU is given by the ICS in 2002 as "a contraction of reduced strength and/or duration, resulting in prolonged bladder emptying and/or a failure to achieve complete bladder emptying in a normal time span" [4]. DU is generally understood to be either the impairment of detrusor muscle contraction resulting from neuronal damage, or changes to the detrusor resulting in impaired contractile function [16]. This loss of detrusor functionality is also referred to in the literature as impaired detrusor contractility (IDC). This has been described as having two fundamental characteristics; the strength of the contractile force and the ability to adequately maintain these contractions [17]. IDC and DU are sometimes used interchangeably but should not be considered as synonymous, as DU describes a clinical syndrome, whereas IDC describes an attribute of detrusor function [18]. Other terms which describe loss of detrusor muscle functionality include *detrusor areflexia*, *hypotonic bladder*, *detrusor failure* and *bladder failure*. These all have their limitations and reflect the past ambiguity of terminology [15]. Even the ICS definition of DU does not define reduced strength or duration, prolonged bladder emptying, completeness of bladder emptying, or a 'normal' time span. This lack of consensus of terminology and complex aetiologies that result in DU are some of the barriers to having standardised parameters that define urodynamic DU [18–20]. However, studies which patients with LUTS referred for further urodynamic investigation, have used various criteria for urodynamic DU which may be considered. Most recently, Gammie et al. [21] and Uren et al. [22] use a bladder contractility index (BCI) of <100 and a bladder outlet obstruction index (BOOI) of <20 in men. This combination of BCI with BOOI thresholds restricts the acceptable degree of BOO, resulting in the exclusion of patients with probable co-existing DU with BOO (those with a low BCI yet relatively high BOOI). Women were required to have a detrusor pressure at maximum flow ($P_{det}Q_{max}$) of <20 cmH_2O and a maximum flow rate (Q_{max}) of <15ml/s. These thresholds were later endorsed by Fode and Sønksen [23]. However, these

parameters are not likely to be definitive, given the ongoing research efforts of many, and further complications introduced through the proposed classification by aetiology [24].

Going forward from this non-specific definition of urodynamic DU, a working symptomatic definition for UAB was derived by a consensus group at the 2014 International Consultation on Incontinence – Research Society as follows; "*The underactive bladder is a symptom complex suggestive of detrusor underactivity and is usually characterised by prolonged urination time with or without sensation of incomplete bladder emptying, usually with hesitancy, reduced sensation on filling, and a slow stream*" [14]. This definition of UAB raises some important points for discussion. Firstly, the underactive bladder is the symptom complex *suggestive* of detrusor underactivity, so UAB is not considered synonymous with DU. The implication is that that a diagnosis of DU on urodynamics is not essential for a patient to have UAB, just that the symptoms in the definition are presented. The underlying aetiology and pathophysiology are also not specified, so UAB could also be comprised of a highly heterogeneous group. For example, patients may have impaired contractility with other urological conditions such as DO, BOO or urodynamic stress incontinence, and still be classified as having UAB. Furthermore, the overlap of LUTS associated with these conditions, such as a slow flow, high urinary frequency and incontinence [14,15,25], presents a further challenge to deriving a symptomatic definition. The need for further qualitative and quantitative research to elucidate the patient reported symptoms on which a symptomatic definition may be based is increasingly recognised [14,26].

The latest symptomatic definition proposed by a Working Group set up by the ICS in 2016 is: "*Underactive bladder is characterised by a slow urinary stream, hesitancy and straining to void, with or without a feeling of incomplete bladder emptying and dribbling, often with storage symptoms*" [27]. The ICS Working Group gives some explanatory notes. Firstly, that UAB "*occurs in association with diverse pathologies and based on current knowledge there is no single distinguishing symptom*". This is in contrast to OAB, for which urgency is the central defining symptom. Secondly, that "*storage symptoms in UAB are varied and may be highly prevalent, including nocturia, increased daytime frequency, reduced sensation of filling, and incontinence*". Finally, "*underlying mechanisms of storage symptoms are diverse and are often related to a significant post voiding residual urine volume*". Other latterly proposed definitions recognise the overlap of possible co-existing bladder outlet obstruction. Fode and Sønksen give theirs as: "*Underactive bladder is the subjective feeling of prolonged urination time, slow stream, and hesitancy, which may or may not be associated with poor bladder emptying and subsequent storage symptoms in men and women without evidence of any outlet obstruction*" [23]. Dewulf et al [24] supplement the original definition by Chapple et al. (2015) using symptoms

shown to be more common in patients with DU than in patients with BOO [21]: "Underactive bladder is a symptom complex suggestive of DU and is usually characterized by prolonged urination time with or without a sensation of incomplete bladder emptying, usually with hesitancy, reduced sensation on filling, slow stream, *palpable bladder, always straining to void, enuresis, and/or stress incontinence."* . Here, the concept of UAB is therefore a term which encompasses LUTS related to impaired bladder emptying, which cannot be primarily attributed to the presence of BOO.

Crucially, the terminology used must meet the needs of both patients and healthcare professionals. It is a specific expectation that both professionals and their patients equally understand the descriptive terms used for their condition. Underactive bladder as a concept is a comparatively simple term, and one that is reasonably straightforward for patients to take on board [2]. However, the complications of deriving a specific symptomatic definition are apparent. Even for patients, as the terminology does not make the contrasting storage and voiding functions clear, the possibility of having both an underactive and an overactive bladder may be intuitively confusing. In conclusion, since there is no normative data or clear cut thresholds, the term for UAB is vague in its description, and rather discursive. Healthcare professionals can be rather dismissive of the term as a result. Nevertheless, conceptually, the term is considered useful for both clinical and research purposes, and it is absolutely certain some patients do manifest weakness of their bladder contractility when attempting to void [2]. Whilst there are differences between proposed definitions, there is also a core consensus which intuitively fits with the concept of an underactive bladder.

1.6 Epidemiology of underactive bladder

The lack of consensus on definitions of urodynamic DU, the similarity of LUTS from often co-existing urological conditions, and that DU is fundamentally a urodynamic observation, means the extent of the underlying contribution of DU to the prevalence of LUTS in the general population is difficult to determine [28]. However, some studies have attempted to investigate the prevalence of underactive bladder using symptoms alone. Valente et al. (2014) evaluated responses to an 18-item survey, mailed to 5000 randomly selected community-dwelling residents aged 60 years and above in Detroit [29]. Data was collected relating to clinical urinary symptoms and demographic characteristics. From a total of 633 respondents to the survey, 23% reported difficulty emptying their bladder and 48% needing to strain to empty their bladder. The authors conclude on the basis of this evidence that DU is 'common and morbid'. However, there is an inability to distinguish DU based on clinical symptoms from other (often

co-existing) disorders such as DO and BOO. Without a urodynamic diagnosis the contribution of DU to the symptoms reported within this study are impossible to establish. The low responder rate is also is also concerning due to the possible bias caused by those with symptoms being more likely to respond. Nevertheless, a relatively small number of studies have used urodynamics to estimate the prevalence of DU in patients who present with non-neurogenic LUTS, albeit with variable definitions of urodynamic DU. Table 1 gives the details of the sample size, diagnostic criteria and prevalence of underactive detrusor found in these studies. These may be summarised as DU being present in 9-10% of men less than 50 years of age, and up to 48% of men older than 75 years of age. Prevalence is less in women at 12-19%.

Table 1. Detrusor underactivity prevalence in men and women when presenting with LUTS.

Study	Sample size with DU (n)	Age (years)	DU urodynamic diagnostic criteria	Prevalence (%)
Abarbanel et al. (2007) [30]	Male: 82 Female: 99	≥70	Q max<10ml/s $P_{det}Q_{max}$ <30 cmH_20	48 12
Jeong et al. (2012) [31]	Male: 632 Female: 547	>65	Male: BCI<100 Female: $P_{det}Q_{max}$ ≤10 cmH_20 and Q_{max} ≤12ml/s	40.2 13.3
Hoag and Gani (2015) [32]	Male: 25 Female: 54	Mean: 59.2 (range 19-90)	BCI<100 and absence of identifiable BOO	23
Fusco et al. (2001) [33]	Men 541	Mean 64 (range 26-89)	$P_{det}Q_{max}$≤30cmH_2O Q_{max}≤12ml/s	20
Groutz et al. (1999) [34]	Female 206	63	<12ml/s and PVR of >150ml	19.4
Valentini et al. (2011) [35]	Female 449	≥55	"impaired detrusor contraction leading to prolonged voiding time and high residual volume"	14
Wang et al. (2003) [36]	Men 90	18-50	$P_{det}Q_{max}$<30cmH_2O and Q_{max}<15ml/s	10
Nitti et al. (2002) [37]	Men 85	18-45	BOOI<20 and Q_{max}<12ml/s	9

Abbreviations: Detrusor pressure at maximum flow ($P_{det}Q_{max}$), maximum flow rate (Q_{max}), bladder contractility index (BCI) calculated by BCI=$P_{det}Q_{max}$+5Q_{max}, Bladder Outlet Obstruction Index (BOOI) calculated by BOOI=$P_{det}Q_{max}$-2Q_{max}. Bladder Voiding Efficiency (BVE) = (voided volume/total bladder capacity)×100.

1.6.1 DU and other co-existing urological conditions

Studies show the complexities of the underlying pathophysiology that urodynamics can reveal in patients that present with LUTS. Urodynamic diagnoses such as DU, DO, BOO, urodynamic stress incontinence, low bladder compliance, acontractile detrusor, bladder hyposensitivity or hypersensitivity may be present in isolation or as concomitant conditions [14,33,37]. Joeng et al (2012) evaluated the urodynamic data of 1179 patients aged over 65 years who presented with LUTS [31]. Half the men and three quarters of the women with DU also had other co-existing conditions such as DO, BOO, or urodynamic stress urinary incontinence. An investigation of community-dwelling elderly (≥70 years) men (n=82) and women (n=99) by Abarbanel and Marcus (2007) observed 'impaired detrusor contractility' (IDC) in 48% of men and 12% of women [30]. Of those with ICD, 66% of the men, and 50% of the women also had involuntary detrusor contractions (DO). BOO was found in 40% of the men, and of those who had IDC, 10% also had co-existing BOO. They conclude that by using presenting LUTS alone, it is not possible to determine the underlying pathophysiology contributing to the symptoms.

1.6.2 Gender

Although the prevalence by gender is poorly understood, the studies using referred populations for urodynamics and by symptomatic presentation have found that DU tends to be less prevalent in women than in men [30,31,34,35]. Of the studied patients in Jeong et al (2012), 40.2% of men and 13.3% of women met their urodynamic inclusion criteria for DU [31]. The authors give no explanation for why a higher prevalence was found in men, other than it is likely to be multifactorial. However, by using a referred sample any epidemiological conclusions are likely to be problematic. For, example, there is a greater necessity for diagnostic pressure flow studies performed prior to prostatectomy to differentiate BOO from DU in men [38]. Patients with DU may have less improvement in post-operative outcomes after prostatectomy than those with BOO and no impaired detrusor function, so establishing DU preoperatively is valuable [39]. Thus, males with suspected BOO or DU are recommended to be referred for pressure flow studies so the clinician and patient will have a fully informed choice on whether to proceed with corrective surgery. Referral may also bias the sample in favour of females, for example, in the study by Gammie et al. (2016), of the 437 patients who met their criteria for DU, 70% were women [21] and of the 28,282 patient records reviewed in their database analysis, 76% of the patients were female.

1.6.3 Age

Although DU occurs in men and women of all ages, a relationship of reduced detrusor contractility [40] or increased prevalence of DU with age has been found by several studies [15,20,21,30]. In their study of 1179 patients aged over 65 years who presented with LUTS, Jeong et al. (2012) found the prevalence of DU increased with age in both men and women, suggesting that DU is an important factor in the pathophysiology of LUTS in the elderly [31]. Valentini et al. stratified their study population of 449 women (>55 years) into three groups 55-64, 65-74 and 75-93 years [35]. Detrusor underactivity defined as 'impaired detrusor contraction leading to prolonged voiding time and high residual volume' predominated in the oldest group, and detrusor pressure decreased whilst post void residual increased with age. However, other studies have differing findings, particularly in relation to detrusor contractility and function [41,42]. In a study of 30 healthy female volunteers, Karam et al. (1997) reported that maximum detrusor pressor did not correlate with age [43]. Madersbacher et al. (1998) had similar results, for both sexes (253 men and 183 women) in a referred population for urodynamics, there were no age-related correlation with maximum detrusor pressure [44]. *In vitro* assessment of human detrusor smooth muscle found no evidence of a decline in contractility or excitability with age [45]. Smith et al. (2010) concludes that it is as yet inconclusive whether impaired detrusor contractility is a cause the of age-related impaired bladder emptying [18]. There are limitations with the cited studies, including variable definitions of impaired detrusor contractility and urodynamic DU. Nevertheless, these data suggest that it is not just the contractile potential of the detrusor muscle but the other aetiological factors of DU that may explain age-related trends.

To summarise, despite the absence of reliable background population data, DU appears to be a relatively common urological condition in patients with LUTS. The prevalence of DU by gender, age, or relative to other urological conditions is unclear, and in the cited studies may be partly determined by the study population. Although the community prevalence of DU is not currently known, a proportion of patients with DU in the background population are asymptomatic and do not present for urologic assessment [46], so the prevalence may be underestimated. This is supported by qualitative evidence that suggests many patients with DU successfully manage their symptoms to minimise impact on their lives [22]. Only by accurately measuring presenting symptoms will it be possible to achieve a more robust picture of their prevalence and impact on the population.

1.7 Symptoms of UAB

There is a current lack of qualitative studies in the literature which explore the patient-reported symptoms and impact of DU or UAB. Qualitative work is called for, in particular to inform the further development of the symptomatic definition of UAB [14,26]. Hoag and Gani (2015) evaluated the presenting symptoms of 79 patients with DU, defined as a BCI<100 in the absence of BOO. It was not clear how the presenting symptoms were evaluated. However, the most common symptoms were urgency (63%), slow stream (61%), straining (57%) and nocturia (48%). The symptoms of UAB may reflect underlying DU, but the isolation of symptoms that may be attributed to DU may be complicated by the presence of additional lower urinary tract dysfunctions [32]. For this reason, Hoag and Gani were not able to attribute the presenting symptoms entirely to DU, in particular as co-existing DO in patients was not excluded from the analysis. Furthermore, the underlying aetiology of some presenting symptoms may also be complex, so ascertaining the underlying mechanism and to what extent symptoms can be attributed to DU may be uncertain. For example, the aetiology of nocturia is complex and age-related [47,48],

In a retrospective large scale analysis of a UK hospital database, Gammie and colleagues (2016) have recently made attempts to identify differences in relative occurrence of the signs and symptoms of patients with urodynamic DU, in comparison with patients with BOO and those with 'normal' pressure flow studies [21]. Many symptoms and medical history factors showed a statistically significant difference in relative occurrence. In male patients, the symptoms of decreased urinary stream (56% in patients with DU, 82% of patients with BOO, and 30% in those with normal PFS) and hesitancy (51% in DU, 69% BOO, 26% with normal PFS 26%) were in high frequency. In women, a decreased urinary stream was present in 29% of DU patients, 20% of BOO patients and 4% of those with normal PFS. However, as symptoms that were specific to a particular group were low in prevalence, and the relative differences in symptom occurrence generally indistinct; information that could be used to symptomatically differentiate DU from BOO still remains unclear [27].

1.8 Impact of UAB

Although there is no UAB-specific qualitative research, it is known that there can be a broad impact on patients' lives through qualitative studies associated with LUTS [9,49,50]. Disruption to sleep due to waking several times in the night and the lifestyle inconveniences caused by increased daytime urinary frequency can be particularly bothersome. The necessity to plan

ahead for awareness of the location of toilets, impairment of social life, embarrassment in particular situations and reduced self-esteem are a feature of qualitative studies in male patients with LUTS [50,51]. However, overall, LUTS do appear to be often tolerated and well-managed by patients [8,50]. Many patients with UAB have a high post void residual and perhaps correspondingly, a high proportion of patients self-catheterise and experience UTIs [32]. The consequences can be very bothersome, in particular the impact of nocturnal voiding, urinary tract infections, and the inconvenience of self-catheterisation on day-to-day life [9].

1.9 Pathophysiology of the underactive bladder

The control of micturition requires complex connections between areas of the central nervous-system (CNS), and involve the sympathetic, parasympathetic and somatic systems [52,53]. The neural pathways between the CNS, urinary bladder and urethral outlet are regulated by simple on-off switching circuits by which voiding and storage functions are achieved. Reflexes that control storage are organised in the spinal cord during bladder filling, whereas voiding reflexes are mediated by parts of the brain [53]. During filling, the parasympathetic detrusor innervation is inhibited, and the urethral sphincter (striated muscle) is activated, which prevents the bladder emptying involuntarily. Voiding is initiated by the relaxation of the urethral smooth muscle and sphincter, followed by the contraction of detrusor smooth muscle a few seconds later, increasing the pressure inside the bladder and resulting in the flow of urine [53,54]. Neurogenic dysfunctions in the efferent or afferent nerves involved in innervating the micturition reflex, in the brain/spinal cord, or myogenic detrusor muscle failure are some of the probable causes of UAB [55]. Ageing, bladder outlet obstruction, neurological disease and autonomic denervation are other possible causes [46].

1.9.1 Neurogenic disorders

Neurogenic DU can be caused by the dysfunction of the central or peripheral nervous system [24]. DU may be observed in patients with neurological disease such as multiple sclerosis (MS), stroke, Parkinson's disease, multiple system atrophy, or trauma such as spinal cord injury and spinal tumours [24,56,57]. Following investigation of 106 ischemic stroke patients who underwent urodynamic studies, 15% had DU, and 14% had DO with impaired contractility [58]. MS patients frequently present with both voiding and storage symptoms. Neurogenic DO is the most common urodynamic finding in patients with MS, but a combination of urological conditions often present [59]. In a study of 65 patients who underwent urodynamic studies, Amarenco et al.

(2013) found DU in 6% of cases [60]. The site of the neurologic lesion in the central nervous system may determine the type of bladder dysfunction and symptom presentation [61].

1.9.2 Bladder outlet obstruction

Studies have established that morphological changes in the bladder wall can be seen with BOO [62,63], so long-term BOO may be considered as a potential risk-factor for DU [64]. However, in a study of 196 male and female patients referred for urodynamics who underwent a repeat urodynamic assessment, 10 years after their original assessment, there was no evidence that bladder contractility decreased in those with long-term BOO [65]. The bladder contractility in patients with underactive detrusors also did not change over time. In Abarbanel et al. (2007) 10% of men with BOO had co-existing impaired detrusor contractility, so it is known that although not all men with BOO develop DU, some men with DU have co-existing BOO [30,66]. The aetiology of BOO is different in women due to the lack of a prostate gland, and may be a consequence of complications post-incontinence surgery or conditions such as urethral stricture or prolapse. As BOO is more uncommon in women, it is possible that BOO may have an influence on the relative prevalence of DU in men and women [28] but the relationship between BOO and DU requires further research.

1.9.3 Ischemic bladder disease and autonomic denervation

A possible explanation for the increase in LUTS with age may be atherosclerosis and vascular endothelial dysfunction [67,68]. This has also been shown in animal models, for example, vascular damage can eventually lead to bladder underactivity in rats [69]. In rabbits it has been demonstrated that moderate bladder ischemia may cause DO, whereas severe bladder ischaemia can cause impaired bladder contraction [70]. This has led to the suggestion that DO may progress to DU [71], although there is no convincing evidence of this hypothesis [28]. Hyperglycaemia in diabetic patients is also known to be the main factor in causing the autonomic neuropathy that can cause diabetic voiding dysfunction, and may lead to DU/UAB [72,73].

To summarise, the aetiology of impaired bladder emptying is likely to be complex, and associated with diverse morbidities or pathophysiological mechanisms. For these reasons, the tailoring of any treatment to the individual is likely to be most effective at improving outcomes.

1.10 Non-invasive diagnosis

There would be considerable advantage for patients and clinical practice if it was possible to diagnose DU by non-invasive techniques [2,74]. Indeed, cystometry requires the insertion of a catheter, is time consuming, bothersome for patients, expensive, and has a risk of urinary tract infection, haematuria and urinary retention [75]. The work of Gammie et al. (2016) discussed earlier presents several differences in relative occurrence of symptoms, and suggests that a specific patient reported outcome questionnaire used alongside other non-invasive tests may be useful for diagnosis [21]. Ultrasonographic measurement of detrusor wall thickness may add additional diagnostic information and has been shown to diagnose DU and BOO non-invasively [76]. A recent pilot study of 123 men showed that an ultrasound measurement of the detrusor wall thickness of ≤1.23mm with a bladder capacity >445ml is diagnostic of DU [77]. These parameters can be used to diagnose DU in 100% and exclude DU in 85% of patients. In addition, the use of penile cuff-urodynamics to measure isovolumetric contraction strength is promising [78], although neither this method or ultrasonic measurement have yet been tested in clinical trials [79].

1.11 Treatment

The treatment options for patients with DU are currently rather limited. However, there are a number of conservative measures that can be advised, before pharmaceutical options or possible surgical interventions may be considered. For those who suffer recurrent urinary tract infections as a result of incomplete bladder emptying, bladder drainage by catheter is the usual preferred treatment option. The teaching of intermittent self-catheterisation (ISC) is recommended for those patients with sufficient manual dexterity, visual acuity and cognition [80]. Long-term catheterisation by indwelling or suprapubic catheter is an option for those who are unwilling or unable to perform ISC. For some patients with DU who have a particularly slow stream, or bothersome high urinary frequency or nocturia ISC may also be appropriate [80].

The use of outlet reduction surgical techniques such as Transurethral Resection of the Prostate (TURP) may be used to reduce outlet resistance to facilitate emptying of the bladder in men. However, it is widely accepted that patients with impaired contractility have poorer outcomes following surgery than those with confirmed BOO [81,82]. In a study of 224 men with a diagnosis of DU who had undergone TURP, there were no symptomatic improvements after long-term follow-up (>10 years) [82]. It is important that preoperative urodynamic assessment is carried out for those patients with suspected detrusor underactivity, so an informed decision can be

made whether to proceed with surgery. In women, the transurethral incision of the bladder might be effective in treating patients with DU [83]. However, the risk of incontinence due to sphincter damage reduces its clinical appeal [84].

The multifactorial pathophysiology of UAB and DU complicates the pharmacological treatment of underactive bladder. There are often multiple contributing factors to the symptoms presenting within an individual patient, so even when a component on which an agent can act is identified (e.g. to improve contractility of the detrusor, or decrease urethral resistance), there may not be the desired outcome [85]. In addition, voiding occupied only a small fraction of the time of the day whereas medication will be continuously active, making any side effects particularly problematic. Muscarinic receptor agonists such as bethanechol and carbachol work by activating the detrusor muscle. However, there is no evidence of beneficial effects for patients with DU/UAB, as it is likely that the lack of selectivity to target sites makes the action of the drug ineffective [86]. The reduction of urethral resistance by α-adrenoceptor antagonists are widely used and are sometimes partly effective for the management of urinary retention caused by BOO [87]. This may also improve bladder emptying in DU/UAB patients, and some positive results have been shown for neurogenic patients [88] but as yet the evidence for the efficacy is inconclusive [85]. Other mechanisms of action on detrusor contraction and the simultaneous relaxation of the urethra by prostanoids may show promise [89], but clinical trials are required. There is a need to better understand the mechanisms of voiding difficulties and pathophysiology of those with DU, in order to tailor the drug development accordingly.

There are several electrostimulation techniques which are used to treat UAB. Neurostimulation and neuromodulation techniques work by stimulating the bladder or the detrusor muscle by nerve root stimulation at the spinal cord, transurethral stimulation, pudendal nerve stimulation, or at the sacral nerve [90]. However, these techniques may only be possible in certain patients and are still experimental in some cases. Nevertheless, sacral neuromodulation has had some positive results [91], has been FDA approved for the treatment of underactive bladder, and is clinically available.

1.12 Conclusion

Patients with LUTS referred for urodynamic tests are commonly found to have a weak and poorly sustained detrusor contraction, termed detrusor underactivity. When the presenting symptoms of these patients are reviewed, they encompass storage, voiding and postmicturition LUTS. The proposed symptom complex of UAB has been described by the ICS

Working Group as being characterised by “a slow urinary stream, hesitancy and straining to void, with or without a feeling of incomplete bladder emptying and dribbling, often with storage symptoms”. This definition is a good starting point; however, the differentiation of symptoms that may be attributed to DU in the context of other co-existing urological conditions is currently not possible to determine. There is a need for further qualitative and quantitative research to elucidate the symptoms that can be attributed to DU. Overall, the available treatments for underactive bladder are still unsatisfactory. There is scope for the further development of testing of pharmaceutical agents, but further clinical trials are required to test the efficacy of the existing treatments. In conclusion, there is increasing recognition that there is a requirement to understand better all aspects of epidemiology, aetiology and pathophysiology of the underactive bladder to enable further progress on developing effective treatment options to be made. Indeed, ultrasonic measurement of detrusor wall thickness or penile-cuff urodynamics when used alongside a condition-specific patient reported outcome measure may be promising to aid a non-invasive diagnosis and to help achieve some of these aims.

Chapter 2 Questionnaire design and psychometric evaluation methodology

2.1 Introduction

The overlap of symptoms from different lower urinary tract and voiding disorders can present a challenge for clinicians to ascertain a clinical hypothesis by history taking alone. However, there are a number of ways which a clinician may objectively capture a patient's experience in order to formulate treatment strategies. The urinary diary (includes the frequency volume chart and bladder diary), allows a patient to record and self-report their symptoms as they occur, and is accepted as an integral tool in the initial assessment of LUTS. When completing a bladder diary the patient records information such as fluid intake, voided volume, the time of micturition and other symptoms for a few days before their visit to their clinician [92]. Clinician completed questionnaires are available [93,94], however, these are not widely used as it has been shown that interviewer administration can introduce potential bias when assessing patient responses [95,96]. Patient-centered questionnaires are increasingly recognised as the most important way of reviewing symptoms and their impact from the patient perspective [97,98]. A patient reported outcome (PRO) is defined by the U.S. Food and Drug Administration (FDA) as 'any report of the status of a patient's health condition that comes directly from the patient, without interpretation of the patient's response by a clinician or anyone else' [99]. Over the last twenty years, there has been a large growth in the number and scope of available PRO measures (PROMs) for the assessment of LUTS, and their impact on health-related quality of life (HRQL). PROMs are distinct from patient reported experience measures (PREMs) which are concerned with measuring a particular service or aspect of their care [100].

The development of a new PROM is a detailed and lengthy process involving numerous sub studies [101–104]. The following chapter will give an overview of the rigorous process involved in the design and validation of a new PROM. Chapter 3 goes on to review some the important instruments that are available for use in research and clinical practice for LUTS and evaluate the necessity for the development of a new condition-specific instrument for the assessment of UAB.

2.2 Design and use of PRO measures

Patient reported outcome measures enable the capture of the subjective patient experience in an objective and measureable way. They are recognised as providing valuable information of patient reported symptoms and impact on quality of life, and to allow assessment of change during ongoing clinical evaluation, research or as outcome measures in clinical trials. In clinical practice it is possible to monitor the ongoing progress of a patient, to formulate clinical hypotheses and manage treatment accordingly. PROs may be used for epidemiological research or the monitoring of health services [105]. In clinical trials PROs are often used as primary outcome measures for the measurement of change in symptoms or HRQol aspects relating to the intervention in question [106,107].

PRO measures tend to be structured as a series of simple questions (or items) with fixed responses which patients can understand and answer quickly and easily. The patient is asked to provide information about perceived severity of symptoms or impact on quality of life over a recent period of time known as a 'recall period'. The questions include a series of possible answers (response options) which allow the patient to indicate the frequency or severity by which they experience a particular symptom or impact on quality of life. These response options are usually assigned points from low to high severity, which allows the calculation of an overall symptom score. Through multiple administrations of the measure it is therefore possible to monitor the ongoing progress of a patient, to formulate clinical hypotheses and manage treatment accordingly [103].

There are three main measurement properties a questionnaire must demonstrate. The first two, validity and reliability are traditionally considered the fundamental characteristics that a PRO instrument must possess [103,105,106]. The questionnaire must be shown to be able to measure the symptoms or health-related quality of life aspects in an accurate (valid) and stable (reliable) manner [101,103,104]. The third property, responsiveness, is of importance if a questionnaire is to be used in clinical applications as an outcome measure to detect change in a patient's condition. The concept of validity relates to whether the instrument actually measures the construct that it intends to measure [101–103]. Along with reliability, it is the degree of confidence that may be placed on interpretation of the instrument's measurement or score [104]. One aspect of validity is *content* validity. This refers to whether the PRO instrument covers all relevant content, without omitting any important issues, but also excluding any irrelevant items [101,104,109]. As PROMs are fundamentally reported by the patients, qualitative research methodology is essential to the establishment of content validity [111,112]

Most importantly, content validity must be supported by direct patient input of the relevant clinical population [112].

The first section of this chapter will discuss in detail the qualitative methodology required to demonstrate content validity when developing a new PRO instrument. The chapter goes on to describe methodologies for the quantitative evaluation of an instrument's psychometric measurement properties. In questionnaire development, psychometric evaluation provides further evidence of validity, reliability and responsiveness in order that users may have confidence in the measurements that are made [104,113].

2.3 PRO measure development good practice guidelines

The release of the United States FDA 'Guidance for Industry: Patient Reported Outcome Measures: Use in Medical Product Development to Support Labeling Claims' in 2009 [99] and the subsequent guidance 'Qualification Process for Drug Development Tools' [114] document reflects the increasing importance placed on the use of PROs in the drug development process. These documents along with reports by Patrick et al. by the International Society of Pharmacoeconomics and Outcomes Research (ISPOR) [112,115] provide direction to ensure the methodology used for the development of an outcome measure for use in clinical trials is rigorous and standardised. The following description of the qualitative methodology used in the development of a new PRO instrument is informed, in part, on the recommendations described by these sources. In particular, if a new PRO measure is intended to be used as an outcome measure in clinical trials (e.g. to support labelling claims), then adherence to these regulatory guidelines during development of the instrument is of particular necessity [99,114].

2.4 Context of use

When developing a new PRO measure the first decisions to be made surround the *context of use*. This is defined by the purpose for which the PRO measure is likely to be used, the disease which is to be measured, and the target patient population [112]. It is helpful to develop an *endpoint model* which clearly sets out the intended context of use for the instrument. For instance, the endpoint for a PRO instrument with the intention for use in clinical trials should specify the condition or disease which is to be measured and the demands on the instrument in order for the trial objectives to be fulfilled (e.g. to provide symptom scores of treatment versus comparison groups, or to detect improvement due to an intervention). If an instrument is to be used as an outcome measure in a clinical trial, the endpoint is particularly important to

be considered in order that potential requirements for medical product labelling claims are met [99].

The target population should be considered and the extent to which the condition may vary across ethnicities, cultures or countries. A diverse a sample as possible is required to ensure the eventual instrument represents demographic variations of the potential questionnaire respondents. If the PRO measure is likely to be used internationally then language is likely to be a significant factor for consideration during derivation of item wording. The content validity of the instrument in different countries is supported if concepts are elicited in the population in which they are intended to be used (e.g. by the consideration of concept elicitation interviews in simultaneous sub studies in different target patient populations).

2.5 Literature review and expert clinical panel

The review of existing qualitative literature gives insight into the known patient experience of the condition to inform the development of the conceptual framework. The review of the published properties of these existing questionnaires determines the necessity or justification for the development of a new questionnaire. In addition, it allows the identification of domains or concepts of interest that may relate to the condition of study. The literature review and consultation with clinical experts then inform the development of a *conceptual framework* or *disease model*. The conceptual framework sets out the concepts to be measured by the instrument in a diagram that presents a description of the relationships between items, domains (sub concepts), and concepts to be measured [99]. This organises and documents hypothesised concepts that are likely to be considered for inclusion as items in the instrument. An expert clinical panel made up of multidisciplinary health professionals in the field is consulted on the clinical relevance of these hypothesised concepts. The collation of this information also contributes to the development of an exploratory interview guide for the initial qualitative interviews or focus groups [112,115].

2.6 Qualitative interviews and/or focus groups

The use of individual semi-structured interviews and/or focus groups are the qualitative data collection methods used in concept elicitation for PRO measure development [116]. Focus groups are usually six to ten people who are encouraged to discuss around the topic, allowing the comparison and enrichment of individual experiences by the feedback and social interaction of a group dynamic [112,117]. An experienced group facilitator is essential for focus groups to ensure the discussion is kept on the subject matter, or the tone is not unduly swayed by a particularly

assertive member of the group. The data generated by focus groups can be difficult to analyse due to multiple participants expressing their views, but this can also provide a rich source of data. Individual interviews are better suited to topics which are personal or sensitive in nature or when a discussion requires particularly detailed experiential information [112]. Although data collection may take longer in one-to-one interviews, the process of transcript analysis can be less complex and scheduling of interviews is generally easier. Interviews must be conducted by a trained qualitative researcher with excellent personal communication skills to elicit the important issues.

Leidy and Vernon (2008) describe three types of qualitative interviewing that are recommended in PRO measure development: exploratory, developmental and confirmatory [117]. Exploratory interviews are used to explore around the topic for which little may be understood. This provides better understanding of the condition and the concepts which are likely to be explored in later interviews and provide additional information surrounding the potential endpoint of the PRO measure. There can also be the incorporation of important concepts into the ongoing development of the interview guide. Developmental interviews build on this process by focussing specifically on the language, phrases and expressions that patients use to describe their experience to inform the item content. Here, the endpoint of the PRO instrument drives the objective of the interviews, for instance, if an instrument is intended to assess symptoms and their bother then the interview questions should be designed to elicit this from the patients. Finally, confirmatory interviews can be conducted to document content validity and identify specific concepts, words or phrases that can be mapped on to the new instrument.

2.6.1 Conduct of interviews or focus groups

Concept elicitation interviews or focus groups are audio recorded to allow the transcription of the content for later review and analysis. Audio recordings rather than video recordings are generally more acceptable for patients when talking about sensitive topics and aid with participant anonymity [112]. The confidentiality of the recordings is assured to the patients and full written informed consent is taken for their use. It is important to quality check and clean the interview transcripts to ensure accuracy to the original source, as the dialogue provides essential context which shapes the analysis.

2.7 Interview guide

The interview guide is generated following the literature review and input from the clinical experts. It represents a clear set of questions and probes that are intended to elicit open-ended responses from the interviewees in a semi-structured manner. A well designed interview guide is important in order to help avoid possible interviewer bias, or unintentional influence of the researcher on the type of responses from the participant [118]. Questions should be carefully worded to be open-ended, non-leading and include the concept of interest to ensure the question has the required specificity. If possible a timeframe which is appropriate can be included in questions to provide context and to reduce recall bias. For example, if the concept of interest was back pain a suitable question would be "How did your back feel over the last 24 hours?" Follow up probes may be included which serve to explore the concept in greater detail such as: "How about when you got up in the morning? Or when you went to bed?". These probes should be used after a patient has been encouraged to speak spontaneously about their experiences. Ideally, a draft guide is tested in the target or similar population before the data collection to identify questions that require improvement or to improve the flow of the interview. Iterative revisions are also made to the schedule in ongoing interviews to follow-up of concepts of interest that may require further exploration.

2.8 Sample size and qualitative data saturation

There is no set number of interviews or sample size required for PRO measure development, as this is dependent on the number of measured concepts elicited, the complexity of how these concepts are experienced by patients and the heterogeneity of the patient population characteristics under study [111]. The current guidance is that interviews are generally scheduled until the concept may be considered *saturated* [112,114]. The simple definition of saturation in the context of qualitative research is the point at where no new information is obtained from additional qualitative data [119]. More specifically in the PRO field, saturation has been defined as *'the point in the data collection process when no concept-relevant information is being elicited from individual interviews, or focus groups, or no new information is deemed missing during cognitive interviewing'* [120]. The purpose of demonstrating saturation in PRO research is to document that the condition under scrutiny has been adequately explored from the perspective of the patient population of interest. Once there are no new important concepts emerging from further interviews, concept saturation can be considered achieved.

Saturation may be documented and demonstrated by the construction of a *saturation table*, which supports the content validity of the tool under development [99,121]. An example of a saturation table for a set of concepts elicited relating to stress urinary incontinence is shown in Table 2 (adapted from [111]). There are no new concepts elicited by interview 3 so this set of concepts may be considered saturated.

Table 2. Saturation table example for stress urinary incontinence.

Stress urinary incontinence concepts	Interview 1	Interview 2	Interview 3
Leakage when sneezing	X	X	X
Leakage when exercising	X		
Leakage when coughing		X	X
Leakage when laughing	X		X

Although the assessment of saturation is useful conceptually, it provides little guidance to accurately predict sample sizes before data collection. However, it is likely that in a relatively homogeneous study population saturation is likely to occur in the first twelve interviews [122]. Within PRO research there is a move towards including a high diversity of populations within clinical trials, so the choice of sample and expected sample size should reflect this. Therefore some contingency should be planned for when predicting sample sizes, as a more diverse sample will mean that saturation may take longer to achieve [123]. Typical sample sizes in recent PRO instrument development studies tend to be thirty or forty interviews [9,124,125].

2.9 Analysis of qualitative data

There are a number of approaches derived from different intellectual traditions which exist for interpretative qualitative analysis. It is beyond the scope of this thesis to explore these methods in detail. However, when selecting the approach to analysis of qualitative data, it is most important that the theoretical framework and methods should be appropriate to the research objectives [126,127]. It also may be argued that there should be flexibility in the approach in order to answer a range of possible research questions [128]. Smith and Firth [129] describe three general methods of approach to the analysis of qualitative data which are commonly used:

- Sociolinguist methods (e.g. discourse and conversation analysis which look to understand the meaning of language [130]).

- Methods which use inductive logic to develop theory (inspired by grounded theory [131]).
- Interpretive methods to describe participants' experiences (e.g. content analysis [132] and thematic analysis [126]).

In PRO measure development, there are no specific guidelines to the approach. This is often the case when employing qualitative research methodology, as there is considerable overlap of the techniques and methodologies used. However, methodology for qualitative analysis both in and outside PRO research are often inspired by grounded theory [133]. Here, the development of concepts or theory evolve 'through continuous interplay between analysis and data collection' [131]. Grounded theory incorporates three types of coding [130,133]:

- Open coding (where the data is examined, compared and categorised by concepts),
- Axial coding (the data is re-categorised based on the relationships or connections between categories)
- Selective coding (the core phenomenon is identified that represents the whole dataset and data is selectively coded that relates to this)

The initial approach to coding interviews for PRO research recommended by Patrick et al. (2011) [112] lends itself well to general inductive coding [134]. The ideas generated are not unduly influenced by prior knowledge and are wholly derived from the patient data. Resulting themes are developed 'from the field upwards' rather than having preconceived ideas [135]. This type of analysis is suitable for topics where very little is known, and allows the documentation and naming of codes in the patient's language. This is appropriate when devising items to maximise patient comprehension at a later stage. However, during PRO instrument development it is unrealistic to expect the researcher to restrict oneself to an entirely inductive theoretical commitment. It is acknowledged that the naming of codes may, and perhaps should, be influenced by current theory (e.g. recognised symptoms). Kerr et al. [121] cites the technique of thematic analysis [126] as a good balance between inductive and deductive logic which may be suitable for the specific objectives of analysis required by PRO research. In essence, within PRO measure development, the coding requires the organized cataloguing of the patient's experience within specific context set by the population of interest and vision of the PRO to be developed.

2.9.1 Coding process

During familiarization of the transcripts, the data is read and assessed to define and name new codes. Newly collected data is compared with previous data according to the principles of the

constant comparison method (rooted in the principles of grounded theory [133]). The accuracy of the codes are re-evaluated as the interviews progress, and modified according to the patient's descriptions. Where possible, the names of the codes are based on the wording of the patients, as they represent not the researcher but the patient perspective. This process of collection of new data through interviewing followed by transcription, coding and evaluation and modification of existing codes is inherently iterative in nature. The eventual endpoint is a final coding dictionary, with each code defined and presented in a way which exhibits how each code interrelates within the context of the patient quotes.

Qualitative data analysis programs such as Atlas.ti [136] or NVivo [137] can facilitate the process of categorising and organizing the data. These programs do not aid with the actual coding; the skill of assigning the right code to the patient expression is dependent on the decisions of the researcher.

2.9.2 Rigour in qualitative analysis

As with any qualitative research, when analyzing data there is an onus on the researcher to reflect on their own perspective and to recognise that their own decisions are influenced by their prior knowledge and experiences. This is known as the principle of reflexivity and is fundamental to the credibility of qualitative research [127(p20)]. Through continual reflexive critical analysis the researcher strives for 'objectivity' in order to maximise the overall credibility [138].

In order to introduce further rigour to the coding process, good practice in the analyses of qualitative data recommends the use of multiple coders [112]. A qualified qualitative researcher not affiliated with the project is given a selection of interviews for review and coding. An advantage of double coding is that an independent coder not immersed in the project may perceive different items of information that are worthy of scrutiny. This provides multiple perspectives to the analysis of the dataset, as it is acknowledged that there will be different ways of categorising the same data. Coders will independently complete the analysis of one or two transcripts at regular intervals throughout the data collection period. Meetings between the coders are arranged to evaluate and reconcile areas of inconsistency between the coding framework and dictionary. In addition, this process evaluates the interviewer consensus regarding the meaning of the concepts, determines whether the major concepts of the dataset have been coded, and that there are no significant omissions.

The process of inter-rater agreement (IRA) may be used to illustrate the stability, accuracy and reproducibility of a coding scheme [139]. Correlations of 0.80 or higher between raters

demonstrate good inter-rater reliability. This level of quantitative agreement evaluation is not usually necessary for self-administered questionnaires, but may be required for instruments using observer ratings or multiple interviewers [111].

2.10 Comment on FDA guidance to qualitative approach

As already mentioned in this chapter, there is a considerable importance placed on the FDA guidance by PRO researchers and pharmaceutical sponsors, due to the significance of being able to support a labelling claim based on the eventual PRO measurements. However, there is little detail on the type of sampling and no recommendations on the approach to qualitative analysis. The FDA guidance states that documentation must be provided to demonstrate that saturation has been reached. In addition, it recognises that a sample size cannot be recommended but the sample should reflect 'variations in severity and in population characteristics such as age, sex, ethnicity, and language groups in accordance with the anticipated clinical trial design'. Arguably, this is advantageous as it allows the PRO researcher greater flexibility. However, Kerr et al. (2010) [121] argue that there is a conflict here; whilst there is an attempt to increase rigour by demonstrating concept saturation, the sampling strategy is at theoretical odds with this. The advised sampling strategy would suggest a quota or maximum diversity sampling should be used to maximise the generalisability or representativeness to the intended clinical trial population. However, the concept of interviewing to saturation implies an approach linked with grounded theory, and thus a theoretical sampling approach.

Kerr and colleagues also question whether data saturation tables are meaningful. Effectively, the researcher makes a choice whether to assess saturation at the level of broad themes or at a high level of detail. It is easier to demonstrate saturation if only the broad themes are listed as opposed to one which details a higher level of analysis. The attempt at demonstration of rigour must not be at the expense of the 'subjectivity and creativity necessary to develop a meaningful understanding through qualitative inquiry' [121,140].

Three recommendations are made for the study protocol [121]. Firstly, to incorporate an element of iterative data collection and analysis which allows the saturation of the data to be assessed whilst data collection is ongoing. Secondly to have an existing planned procedure to assess saturation, and finally, to allow flexibility to have further interviews to achieve saturation if required.

2.11 Generation of draft PRO instrument

2.11.1 Item generation

The selection of items to include in the PRO instrument is mainly related to whether a concept is mentioned frequently by patients in the concept elicitation interviews. However, an item may also be considered for inclusion even if mentioned infrequently, for example, if it was mentioned by only 5% of patients but was described as a particularly bothersome and important symptom by this minority. Whether or not the concept was mentioned spontaneously (without prompting) in the interviews can also add weight to the decision to include a particular item [141]. Expert clinical input can be useful to provide insight into the diagnostic value of including an item, or mechanisms which contribute to an illness [101,142]. The published literature should also be used to determine which concepts to include as items. Using this information, the instrument developer makes decisions on the criteria which are suitable to evaluate the specific condition within the context of measurement. The overall consideration is to include items that capture the range of important concepts that represent the experience of the condition which is to be measured [115]

The wording of items will be dependent on intent of the item, expert clinical input and the qualitative data obtained in the concept elicitation interviews. The PRO instrument should use terminology that is familiar to the patients who will be reading the questionnaire. For example, the use of the term 'somnolence' in an item may be appropriate and valid in a clinical setting but may not be understood by the patient population. The use of 'sleepiness' or 'drowsiness' in the item stem may be more appropriate. The way the concepts are described by patients in the concept elicitation interviews is a guide to the words which are used in the developing instrument.

2.11.2 Recall period

The period of time over which the patient is asked to recall their symptoms is an important consideration. A recall period that is too long may run the risk of introducing recall bias, where the respondent is not able to accurately remember their symptoms, but too short, and the necessary information may not be captured. Multiple administrations may also be required which increase the burden on the patient. The FDA recommends a recall period which is as short as possible in order to minimise recall bias, whilst also taking into account potential respondent burden [99]. The selection of an appropriate recall period will depend on the purpose of the instrument, the frequency of assessment and length of time over which the condition of

interest provides a stable measurement [143]. Patients will also describe the length of time they observe symptom or impact fluctuations which can influence the sensible choice of recall period.

2.11.3 Response scale

The selection of a response scale is an important aspect of the item design that should occur in conjunction with development of the question items. Firstly, the developer must choose whether to measure severity, frequency, or length of time in relation to a concept. For example, an item measuring pain could evaluate how severe that pain was, the number of times the pain was experienced, or the duration for which the pain has been experienced. There are also a number of different types of scales to choose from, as described in Table 3 [144]. The advantages or disadvantages of using a particular response scale will depend on the type of concept which is to be measured.

Table 3 Examples of response scales. Adapted from (Hambleton et al., 1991)

Example 1. Numerical rating scale (average severity) Please rate your pain by circling the number that best describes your pain on average in the past 24 hours 0 1 2 3 4 5 6 7 8 9 10 No pain at all — Pain as bad as you can imagine
Example 2. Visual analog scale (peak severity) Please mark a (/) to indicate how you have felt over the last 24 hours Not at all ---------- Extremely
Example 3. Likert-type scale (patient impression of change) Since starting the medication, my knee pain is: Very much improved Much improved Minimally improved Unchanged Minimally worse Much worse Very much worse
Example 4. Categorical scale (occurrence of waking) Did you wake up in the night due to shortness of breath? YES NO
Example 5. Numeric scale (frequency of headaches) How many headaches did you have in the last week? Enter number _____ headaches
Example 6. Amount of time (duration) Thinking about the migraine headache that you had today. How long did it last? Enter amount of time_____ hours(s) and _____ minute(s)

2.11.4 Order, formatting and layout

Following item generation, the presentation, order and organisation of the instrument must be considered. Often, items which cover similar concepts are grouped together in 'domains' to facilitate the ease of completion for the respondent. The order of items may be important to consider if they are required to be answered sequentially, have similar response scales, or a subsequent question is dependent on the answer of the preceding item for example. Decisions are to be made on the presentation, including the use of tick boxes, circles or numbers; boxes to clearly separate out items; instructions to skip questions; page breaks, and font size or type [115]. Indeed, a simple improvement such as increasing the font size and having thicker paper may significantly increase the response rate and completeness by respondents [145]. If the instrument is likely to be administered in electronic format then this may also influence the way items are presented. For instance, items may be presented on a single page in order that respondents only see one item at a time before moving on to the next item. Modifications from an original paper administered instrument when changing to electronic mode may result in the electronic version having to be validated separately to ensure the instrument is still understood as expected [146].

2.12 Cognitive interviews

Cognitive interviews follow the concept elicitation phase and generation of the draft instrument to be tested. There are two primary objectives, firstly to ensure that the PRO instrument content is representative of the most important concepts of interest to the studied population within the intended context of use. Secondly, to assess the respondents' understanding of the instrument including the facility of completion, their comprehension of the items and response scales and appropriateness of the chosen recall period. The evaluation and documentation of patient understanding of the PRO measure by cognitive interviews provides evidence for content validity [115]. One aspect of content validity is 'face validity'. This is essentially whether the items appear on the surface to be measuring what they actually intend [104]. If items appear to be irrelevant to users the respondents may not answer the item, (irrespective of whether the item is psychometrically robust). Face validity is addressed by essentially asking the intended respondents to provide feedback on the relevance of the item, and can be part of the cognitive debriefing process.

2.12.1 Cognitive debriefing study sample

The number of cognitive interviews required will be variable for different studies and will depend on the complexity of the questionnaire, the concepts measured, and the diversity of the sample population [115,141]. A sample of seven to ten interviews has been suggested to confirm patient understanding of an item [141]. Other studies complete successive rounds of 3-5 interviews whilst making iterative improvements after each round [124]. The main requirement is that subjects recruited for the cognitive interviews should represent the target population in which the PRO instrument will be used (e.g. age, sex, ethnicity, socioeconomic status, literacy and characteristics of condition), so the overall feedback from the participants will be representative of the population of interest. Any participants which may have unique or different requirements when completing the instrument (e.g. visually impaired, low reading ability) can also be purposively sampled to capture their perspectives.

2.12.2 Cognitive interview process

The standard cognitive theory model on which cognitive interviewing is based consists of four stages to explain how items are processed and answered by subjects [147]. Respondents must first understand the question, then recall the item-specific information, assess the type of information required and finally decide on the response. During cognitive interviews patients are asked to read an item, then asked a series of questions by the interviewer exploring how they interpreted the item and what it meant to them.

The cognitive interview process is based on a 'thinking aloud' approach [148], where the subject is encouraged to verbalise their thought processes as they complete the questionnaire. A semi-structured interview guide is a useful tool for the interviewer to guide the subject through a series of systematic questions which direct the process. A useful starting question can be a simple question such as 'can you tell me in your own words what you think this question is asking you about?' This is then followed up by more specific probes to ask about the comprehension of other aspects of the items, such the response options, recall period, and format of the item. Essentially, the interviewer is checking whether the subject is interpreting all aspects of the item as expected. If there are any inconsistencies, ambiguities or difficulties in comprehension then the interviewer may ask if the subject has any suggestions for changing the question so that it is easier to complete. Difficulties with understanding of the items may not be verbally communicated so it is important that the interviewer is aware of body language, facial expressions and other signs such as flipping pages back and forth which may indicate confusion. Every aspect of the questionnaire is checked in this way, including any

initial instructions, the recall period which is applicable to items, and the format and layout of items [115].

The purpose of a probing question in an interview schedule is to elicit a particular communication. Nevertheless, however standardised the questions are, there are likely to be differences in the way each question is put to the respondent. The ideal scenario would be that the subject provides all the information with the minimum of questioning, however, probes used to direct the interview will invariably be necessary. This opens up the possibility of interviewer bias [103]. Poorly worded questions can lead the subject into affirming responses that the interviewer would like to hear, or cause misunderstandings that obstruct the flow of the interview. In addition, the subject forms an 'inner picture' when deciding how to respond which can be influenced by many factors such as maintaining appearances, poor rapport with the interviewer, and other private restraints on wishing to divulge information. These influences may or may not translate to important differences in interpretation of the results. However, to minimise any potential bias effects, questions should be open ended and carefully worded. For example, when investigating whether or not the questionnaire is an appropriate length (that the subject does not consider it burdensome) [115]:

Poorly worded: *"Is the questionnaire too long? Too short?*

Preferred wording: *"What did you think about the length of time it took to complete the questionnaire?"*

During the interview, notes are taken in response to the key questions for each question on the interview guide. In addition, the interviews are audio recorded and then transcribed verbatim to allow subsequent review of the patient responses and for documentation purposes. After a round of 3-5 interviews the researcher prepares a summary of the responses of the subjects for each item or line of questioning, including key quotations that represent particularly relevant views on items or concepts.

Based on the feedback from the interviews, decisions are then made to revise items or other aspects of the instruments. If four out of five of the subjects in a round of interviews found a particular item confusing, then it is quite clear that the item may need revision or removal. However, it is less clear if only one subject has difficulty; here the researcher must make the decision whether to modify the item, which may not be straightforward. Common reasons for modification of an item is unintended ambiguity, misinterpretation and lack of clarity [115]. The end goal is to reach saturation of the cognitive evaluation of the instrument. This is achieved

when all items are fully understood or interpreted as expected and the questionnaire is easily completed by several successive subjects.

The results of cognitive interviews may be presented in an item tracking matrix. Information that is included details the initial wording of the item, and each modification made after each round of interviews are included, along with patient quotes which provide evidence to support why modifications were made. Items which were removed as a result of the cognitive interviewing process are also included with the rationale provided. This evidence is part of documenting content validity of the instrument [115].

Following the cognitive interview process the end result should be an instrument which is easily completed and understood by the target population. The instrument is then ready for the next phase of development and to undergo the psychometric testing of its measurement properties.

2.13 Psychometric evaluation methods

Psychometrics in questionnaire development is the scientific method which provides evidence of these capabilities in order that users may have confidence in the measurements made [104,113]. The information collated by patient-reported questionnaires is inherently subjective in nature. Psychometrics in this context is the standardised methodology used to quantify and lend objectivity to the evaluation of their measurement properties. Generally, this is evaluated by another sub-study, with a further sample of the target population. Respondents are asked to complete the instrument to be evaluated and amongst other comparisons, the responses to the items are compared with the responses given to a repeat administration of the instrument. The following section of the chapter describes the methodology by which aspects of validity, and reliability and responsiveness may be quantitatively evaluated.

2.13.1 Content validity

It has been discussed in detail how the process of concept elicitation, literature review, the consultation of clinical experts in the field and the process of conducting cognitive interviews contribute to content validity. When assessing the psychometric properties of the PRO measure, the following analyses also contribute evidence for content validity.

Missing data - The level of missing data is explored for each item, when the questionnaire is administered to the target population. If items are consistently missed by respondents this may indicate that an item is irrelevant, or difficult to answer. Identification of these poorly

answered items allows the items to be flagged for possible modification or consideration for removal, in order that higher rates of completion may be achieved. A level of missing data of 3-5% in PRO items is generally considered acceptable [149,150].

Floor and ceiling effects - The presence of floor or ceiling effects within instrument items can adversely affect the range which an instrument is able to measure [151]. Items are considered to have a floor or ceiling effect if a relatively high proportion of respondents give answers in the lowest or highest response option respectively. Ideally, a well-designed item should elicit responses with an even spread across all response options when administered to potential respondents, in order to maximise the item's potential to detect sensitivity to change in a population. The identification of these effects allows the potential modification of items, and thus, optimise the assessment of the patient's condition [152].

2.13.2 Construct validity

Construct validity is the aspect of validity that provides confirmation that the instrument is measuring the underlying concept that it intends to measure, by comparison with known theory [101]. Hypotheses of how the instrument should 'behave' when compared with expected relationships according to known theory are explored. Construct validity is supported when the PRO instrument is shown to measure constructs that are consistent with these hypotheses [108]. The demonstration of construct validity is particularly important for PRO instruments, as often these aim to measure attributes that are reported by patients and cannot be directly observed (e.g. symptoms, emotions, psychological effects). For example, based on known theory, stress urinary incontinence is known to be more prevalent in women [153], so an instrument should be able to detect this. However, construct validity only lends weight to the validity of the instrument if both the theory and the instrument are correct. It is important that the theory is well evidenced, or the apparent failure of an instrument to detect a relationship may be not due to the instrument but the inadequacy of the theory [101].

There are two fundamental types of construct validity. Convergent validity is the extent to which the instrument correlates with other ways of measuring the same construct. For example, a measure demonstrates incontinence is more prevalent in women. Conversely, discriminant validity is when a measure is shown not to correlate with a measure which purports to measure dissimilar variables [101,108]. For example, an instrument should not demonstrate that incontinence is more prevalent in men when all other theory suggests otherwise.

Construct validity is usually demonstrated by correlating subgroups of the sample (e.g. by gender or age) with hypotheses of how the instrument should perform for these groups. However, ideally, a number of sub-studies are performed that evaluate the PRO measure against other ways of observing the construct (e.g. clinical measurement, objective performance tasks). The rationale here is that the score obtained by the PRO measure under evaluation is not just determined by the concept being measured, but also by aspects of the instrument itself [104].

2.13.3 Criterion (concurrent) validity

Criterion validity refers to the correlation with a 'criterion' scale. This is a measure that may be considered a 'gold standard' by which the PRO measure to be evaluated is compared. However, when a new PRO is under development, there is often no existing gold standard, so other existing PRO measures which measure similar constructs are used for comparison. *Concurrent* validity is a type of criterion validity, and is when an instrument is compared with an existing validated measure that measures the same attributes. The PRO measures to be compared are usually administered at the same time, so 'concurrently' [104].

2.14 Reliability

An instrument will also require testing for its reliability, that is, the ability of the measure to perform in a stable, reproducible and consistent manner [101,103]. The following discusses two methods which provide evidence of reliability; internal consistency and test-retest reliability.

2.14.1 Internal consistency

Internal consistency is the extent of the homogeneity of the items, or the degree to which the questionnaire measures the same concepts [103]. Correlations are performed within the items of the scale, or within groups of items (domains) to explore how related (consistent) the items are to one another. The calculation of Cronbach's alpha (α) gives a statistical indication of this, by returning a value of between 0 and 1. A value of zero means that the items are entirely unrelated to one another, and a value close to 1 would indicate that the items were measuring almost exactly the same attribute. However, care should be taken when interpreting α, as the value is not only dependent on the correlation of the items, but also the number of items. The larger the number of items the more likely the instrument will have a high α. If the number of items or α is high, then this suggests that a number of items are asking about the same concept in slightly different ways, so probably there is redundancy and scope for the removal

of items [103]. Thus, a Cronbach's α of ≥0.7 is generally considered acceptable and >0.9 to indicate redundancy within the item pool [154].

2.14.2 Test-retest reliability

The test-retest reliability gives an indication of the reproducibility or stability of the instrument. It is evaluated by individual response consistency between repeat administrations of the instrument over a time-frame in which the responses are not expected to change [108]. The scores obtained from an instrument (X) consist of two components:

$$X = T_s + e_s$$

where T_s is the hypothetical 'true' measurement and e_s is the random 'error' [155]. A reliable measurement depends on maximising the true score and minimising the error score. The error consists of an unsystematic variation (due to random errors when judgements are made by humans) which affects the accuracy of the instrument, but do not result in an overall bias [156]. For this reason, one cannot expect perfect reliability [103], but an acceptable level of reliability, the criterion of which is determined by the selection of the appropriate statistical test.

The Pearson product correlation coefficient can be used to give a measure of the strength of a linear relationship between the two administrations. However, this may be misleading as these correlations are not sensitive to systematic variation in scores due to unknown differences over time. For example, the second administration may have consistently and artificially inflated scores due to a 'learning effect', or some other external factor or bias which have not been accounted for. Thus, the intraclass correlation coefficient (ICC) is increasingly used as a way of assessing test-retest reliability as it takes systemic variation in scores and variability between groups into account [155,156]. The ICC ranges from 0-1 with the generally accepted criterion for adequate reliability as ≥0.7, and >0.9 as excellent [154].

When evaluating pairs of scores for individual items derived from test-retest data, there is a high chance that the scores will agree by chance, especially if the response options are ordinal (e.g. mostly agree, completely agree etc.) or dichotomous (e.g. yes/no). The kappa coefficient is therefore an appropriate statistical test when comparing pairs of scores for individual items, as it provides a chance-corrected measure of agreement. The Kappa statistic provides an indication of agreement from 0-1 where 1 is perfect agreement whereas 0 is no better than it

would be by chance. (Kappa statistic (κ): 0 = poor, 0.01-0.20 = slight, 0.21-0.40 = fair, 0.41-0.6 = moderate, 0.61-0.8 = substantial, and 0.81-1 = almost perfect agreement [157]).

2.15 Responsiveness or sensitivity to change

If a PRO instrument is intended to be an effective outcome measure for clinical assessment or research purposes, then it must be shown to be sensitive or *responsive* to the change in condition of a patient. Hays and Hadorn (1992) and others [158–160] include responsiveness as one aspect of an instrument's validity. Others argue that it is distinct concept of a measure and separate to validity and reliability, and should be evaluated, if necessary, as a psychometric property in its own right [161,162].

A PRO measure ideally aims to be sensitive to changes in a patient's condition over time. If an effective intervention is administered in the time period between administrations then the scores are expected to change, so the questionnaire must be responsive to any change in the 'true' values of the underlying construct [101]. In order to be sure that any change detected is not due to chance, when testing responsiveness to change the instrument's scores should be correlated with those of another validated measure or treatment administered at the same time. For example, a depression scale might be correlated with the results of a psychiatric interview.

Sensitivity to change can be evaluated using percentage change in total score before and after an intervention of known efficacy [96]. When comparing paired ordinal data, the Wilcoxon matched pairs signed rank test with a Bonferroni correction can be used to establish the statistical significance of any change calculated [160,163].

2.16 Factor analysis

Factor analysis is a useful statistical technique used to identify and define the separate factors that can group an instrument's concepts [101]. It operates on the theory that the measured concepts share underlying latent variables, which are identified by their common variance [164]. Factor analysis is used to find the number of factors which influence the variables and to discover which items may be grouped together. The aim is to find the smallest number of common factors which describe the correlations [165]. It can be used to identify potential items for removal, to group items, determine how strongly items are related, and to provide information that informs scoring [154]. A large sample size of 300 participants with 5-10 observations per variable is recommended, although smaller sample sizes will still yield useful

results [164], and are commonly evident in PRO development studies [124]. Expert clinical opinion should be sought when interpreting the results of factor analysis, as although an item may appear to be weakly grouped and marked for removal, it may have significant clinical value or importance to patients so should be retained. Decisions are made surrounding the inclusion or removal of items in combination with the information surrounding their reliability, validity and sensitivity to change.

2.17 Item response theory

There are two main approaches to evaluating PRO measures [166]: classical test theory (CTT) [167] and item response theory (IRT) [168]. Classical test theory is the traditional and dominant paradigm, and is the conventional approach to evaluating the validity and reliability of an instrument discussed so far in this chapter. Item response theory (IRT) refers to a collection of mathematical models that may be used to explain the relationship between the responses to items and an underlying, unobserved construct [169]. IRT can be used when this unobserved trait may be considered to exist on an underlying continuum, such as the constructs 'fatigue' or 'physical functioning' [170]. The probabilities of the respondent to the questionnaire selecting each of the given response options are expressed as a function of this underlying construct, shown graphically by an item response curve (or item characteristic curve). It is then possible to identify which items are most useful in assessing this construct, and to assess the respondent's own position on the continuum. There are several IRT models which may be selected according to how many parameters and the type of items (e.g. dichotomous or polytomous) which the data has. For example, models that assume unidimensionality (items that measure one concept) include the partial credit model [171], generalised partial credit model, and the graded response model [172].

One example of a questionnaire in which both CTT and IRT was used in its development is the Pain Impact Questionnaire (PIQ-6). This is a six-item measure of pain impact and severity on health-related quality of life [173]. Initially, eight items were selected from an item bank made from items from 16 widely used pain measures. The final six items were selected using factor analysis to assess unidimensionality (whether there was a single underlying construct) and therefore suitability for IRT. The inspection of the item response curves showed that the response categories were 'functioning well for all items'. One item was removed because of 'item misfit' and another because the 'item parameters were almost identical'; essentially they covered too similar concepts. The item response curves also informed the weighting of the scoring (the flatter the slope of the curve the less discriminatory power for severity, so less

weighting is given on the score). Alongside this analysis, traditional CTT psychometric techniques assessed internal consistency, convergent validity, divergent validity and known group validity, providing additional information on the properties of the final questionnaire.

The use of CTT is usually advised for initial descriptive evaluation of the psychometric properties of a new instrument [169]. IRT is often used to provide additional, in-depth information in later stages of psychometric testing, particularly as larger sample sizes of several hundred are required (depending on the numbers of items and response categories) [166]. IRT is particularly useful for informing decision making when at the item reduction stage, as it has the capability of reducing the length of questionnaires without making compromises on reliability or sensitivity [170]. However, both approaches can be used successfully when assessing the content validity of a new PRO measure, with final decisions on the inclusion of items made on the basis of both quantitative and qualitative findings [166].

2.18 Scoring

Most PRO instruments have a straightforward scoring system with each item scored in the same way, with equal weighting. The item scores are then summed up to produce a total score [101]. The issue with the equal weighting of items means that there are multiple ways of achieving the same score using the same items in the questionnaire. Thus, the instrument score itself can only give an indication of the severity of the overall condition, not any detail of the patient outcome. Differential weighting of items can be used to assign higher importance to items. However, this raises questions about how much relative weight should be assigned to each item and contributes to complex scoring algorithms that can be cumbersome for analysis. One method is to assign equal scoring importance to domains of items within the instrument, with each domain consisting of different numbers of items. However, in scales with fairly homogenous items or large numbers of items, the weighting of scoring of items can make little difference in practice [104]. People who score heavily in one section of the questionnaire are likely to score highly in another.

2.19 Translation and cultural adaptation

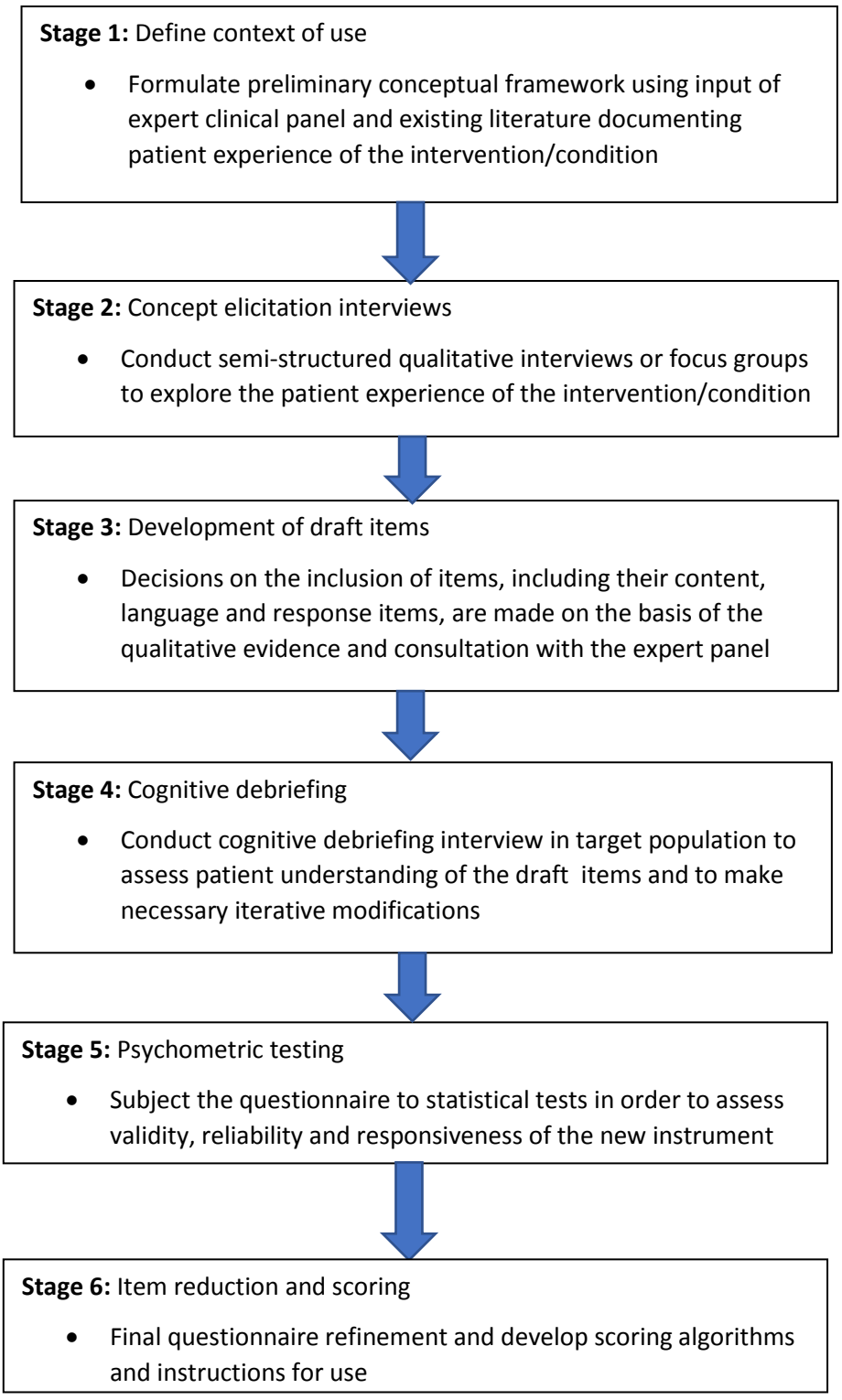
It is often required to translate a PRO into another language from the original. This may be done during the development process to allow the evaluation of its properties in culturally diverse populations, or after the PRO has been finalised to broaden its clinical application. It is important that guidelines are followed when carrying out translational and cultural adaptation to ensure the instrument's conceptual equivalence, cultural relevance, and facility of

understanding [174,175]. This is a rigorous process with multiple steps involved. The main steps are a forward translation into the target language conducted by a native speaker of the target language who is fluent in the original language, followed by a back translation conducted by a native speaker of the original language and fluent in the target language. The back translation is then reviewed and compared with the original and harmonized to achieve consensus. A small number of cognitive interviews with relevant patients who speak the target language are then used to check the understanding and interpretation of the items [174].

2.20 Conclusion

The development of a rigorously validated questionnaire is a lengthy process that involves several sub-studies. The initial drafting of the items follows a literature review, consultation with clinical experts and in-depth interviews or focus groups with patients. Patient involvement in the initial stages of developing a questionnaire is given particular emphasis by regulatory guidelines for industry and good practice. In addition, the psychometric testing of a questionnaire is crucial for the establishment of its fundamental properties of validity, reliability and responsiveness. Figure 1 summarises main stages of PRO measure development described in this chapter.

Figure 1 Summary of stages of PRO instrument development.



Chapter 3 Evaluation of existing measures for LUTS

3.1 Introduction

As discussed in chapter 1, there is a call for a validated condition-specific questionnaire to facilitate the assessment of underactive bladder in clinical practice. The primary purpose of this chapter is to determine the necessity for the development of a new condition-specific measure for UAB. The current literature suggests patients with DU may present with a range of storage, voiding and post-micturition lower urinary tract symptoms (LUTS). It is recognised that there is a significant clinical overlap with many of the symptoms experienced by patients with bladder outlet obstruction (BOO) and overactive bladder (OAB), including high urination frequency, nocturia, urgency, and incontinence [14,25]. PROMs that assess the common signs and symptoms of lower urinary tract dysfunction, BOO and OAB are reviewed for their content and potential relevance to an outcome measure for UAB. Associated quality of life measures are also reviewed in order to explore the range of patient-centered measures available. These are discussed in the context of current regulatory requirements and their published evidence of reliability, validity and sensitivity to change.

3.2 ICI grades of recommendation for questionnaires

The committee 5B for the 6th International Consultation on Incontinence (ICI) 2016 reviewed the available evidence of the psychometric properties for published self-completion questionnaires [97]. The result was a comprehensive publication and review of the measurement properties and status of validity of the published PRO measures relating to the assessment and screening of LUTS, associated HRQL, bother, measures relating to urgency, faecal incontinence and associated sexual function. Each questionnaire reviewed by the ICI received a grade based on the documented evidence of the reliability, validity and responsiveness of the questionnaire as detailed in Table 4.

Table 4. ICI grades of recommendation

Evidence required	Grade of recommendation (+ sign may be added if published evidence of content validity)
Published evidence of validity, reliability and responsiveness to change. 'highly recommended'	A
Published evidence of two of the three main aspects of validity and reliability and responsiveness to change. 'Recommended'	B
Published evidence (including abstracts) indicating the 'potential' for validity, reliability and responsiveness to change.	C

3.3 Current review inclusion and exclusion rationale for existing PROs

For the current review, the content of questionnaires reviewed by the ICI committee [97] were investigated for their potential relevance for the assessment of symptoms of UAB. Of particular interest were those which aimed to measure symptoms and their impact suggested to be associated with a diagnosis of DU or UAB in the literature. Hence, PRO measures which included storage, voiding and post-micturition LUTS and associated HRQOL items were evaluated.

The full text of published evidence of validity was accessed for any PROM that assessed the common signs and symptoms of lower urinary tract dysfunction, including measures for BOO, OAB and associated HRQL. Although questionnaires of grade A or A+ status were of primary interest, questionnaires with a grade of B or lower were also investigated for their potential relevance in content.

In addition, a comprehensive literature search of Medline, Embase, PsycINFO and Google scholar was performed with combinations of the search terms (and associated acronyms): Patient Reported Outcome, Questionnaire, Validation, Lower Urinary Tract Symptoms, Detrusor Underactivity, Underactive Bladder, Overactive Bladder, Benign Prostatic Hyperplasia. The intention was to identify questionnaires of interest that had been developed

since the ICI review, including those which have no evidence of published psychometric testing.

3.3.1 ICI review

Of the 74 measures relating to LUTS and impact on HRQL reviewed by the ICI, 51 questionnaires were Grade B or lower or had a specific focus such as incontinence, pelvic floor distress, patient satisfaction or bowel symptoms. These were not explored further. 23 PROMs assessed the storage and voiding symptoms, signs or HRQL that warranted the further investigation of their content. Upon obtaining the full text for associated publications for of these questionnaires, 12 questionnaires had a specific focus that did not include storage or voiding LUTS (e.g. urinary incontinence, urinary tract infection, post-operative assessment, sexual function) and were not included further in this review. The International Prostate Symptom Score (I-PSS) [176] has a grading of B, which reflects that patients were not involved in its development. Nevertheless, it is widely used in research and clinical practice as an 8-item screening tool for capturing the severity of symptoms relating to BPH, so was reviewed further for its content.

3.3.2 Database search

A search of the databases was carried out to identify publications of the psychometric properties of PROMs relating to LUTS that were not included in the ICI review. Search terms were generated to cover the population of interest, questionnaires and psychometric terminology. Each term were combined with the Boolean operator “AND”. Searches were carried out in February 2015 and then an updated search in October 2017 using the same terms as the initial search. Reference lists of articles were also checked for other potential articles of interest. Abstracts were screened and included studies were in the English language and were a psychometric validation of the properties of a PROM. Studies that were not considered further were those that had patient cohorts that did not have LUTS, or where the questionnaire was not patient reported (e.g. clinician or proxy). Abstracts, conference papers, editorials and review articles were also not considered. The full text of the publications were retrieved for all manuscripts that met these criteria for detailed screening for suitability and final inclusion in the review. Using this process a further 2 questionnaires related to the assessment of LUTS, the LUTSS [177] and LUTS [9,178] tools. The search of Google scholar identified a single questionnaire provided by the Underactive Bladder Foundation, the UAB-q [179]. Figure 2 summarises the process by which the PROMs were selected for review.

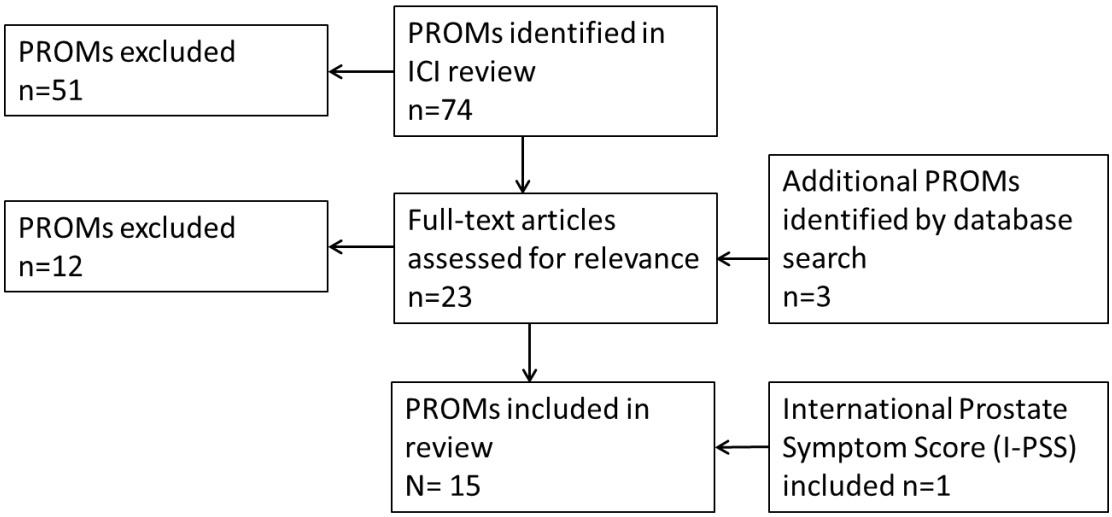


Figure 2. Flow chart of the process by which PROMs were selected for review.

Table 5 describes PRO measures which aim to detect the presence (or lack of) symptoms, their frequency and associated bother. Table 6 describes PRO measures which aim to assess the impact on HRQL. There is a description of the content of the tool and their target population. In addition, their published psychometric properties of validity, reliability and responsiveness are described or indicated.

Table 5. Selected PROs related to the symptom assessment of LUTS. Adapted from Castro-Diaz et al. 2017 [97].

PRO measure name	ICI Grade	Description of Tool	Validity		Reliability		Responsive-ness
			Content	Construct	Internal consistency	Test-retest	
B-SAQ (Bladder Self-Assessment Questionnaire)[180,181]	A	8-item for LUTS for women and validated in men in 2014	Literature review and panel of experts	Concurrent validity. Criterion demonstrated using KHQ	✓	✓	✖
OAB-SS (Overactive Bladder Symptom Score) [182]	A	7-item tool, overactive bladder symptom score for men and women	Initial questions derived from expert panel. Patients involved in cognitive interviews	Discriminant validity compared between diagnostic sub-groups	✓	✓	✖
DAN-PSS-1 (Danish Prostatic Symptom Score) [183,184]	A	12-item tool to assess voiding problems and associated bother in men with LUTS suggestive of BPH	Literature review and cognitive debriefing but not in item generation	Convergent and discriminant validity demonstrated	✓	✓	✓
ICIQ-MLUTS (ICIQ-Male Lower Urinary Tract Symptoms)[185,186]	A	14-item tool to assess male LUTS and BPH with associated bother. Derived from long form ICS-male [186]	In-depth interviews with patients and consultation with expert panel	Compared with pressure-flow studies and correlation with I-PSS	✓	✓	✓
ICIQ-FLUTS (ICIQ-Female Lower Urinary Tract Symptoms) [150,163]	A	13-item tool to assess female LUTS, (particularly UI), and associated bother. Adapted from ICS-male	Consultation of expert panel, literature review and discussion with patients	Compared with sample of asymptomatic men. Criterion demonstrated using frequency-volume diaries	✓	✓	✓
ISS (Incontinence Symptom Severity Index)[107]	A	8-item tool to assess severity of urinary storage and voiding symptoms in women	No evidence of method of item generation	Concurrent and criterion evaluated	✓	✓	✓

PRO measure name	ICI Grade	Description of Tool	Validity		Reliability		Responsive-ness
			Content	Construct	Internal consistency	Test-retest	
OAB-V8 (OAB Awareness Tool) [187,188]	A	8-item tool, aimed as a screening tool for OAB in men and women. Adapted from the OAB-q [188]	Focus groups and a literature review for OAB-q but not specifically for OAB-V8	Concurrent and discriminant evaluated for OAB-q	✓	✓	×
LUTSS (Lower Urinary Tract Symptom Score)[177]	Not graded	14 item tool, LUTS for men and women	Expert panel developed questionnaire from patient interviews followed by subject feedback on drafts.	Discriminant, criterion evaluated using AUASS (I-PSS)	✓	✓	✓
LUTS (Lower Urinary Tract Symptom Tool) [9,178]	Not graded	16-item tool assessing frequency and bother of 18 LUTS in men and women	Focus groups used to explore patient descriptions of LUTS and perspective on treatment outcomes.	No published evidence currently, in development	Not available	Not available	Not available
UAB-q	Not graded	6-item tool assessing UAB	Preface stating developed by experts in bladder research but no published psychometric evidence	×	×	×	×

Table 6. Selected PROs related to the assessment of LUTS on HRQL. Adapted from Castro-Diaz et al. 2017 [97].

PRO name	ICI grade	Description of tool	Target population	Validity		Reliability		Responsiveness
				Content validity	Construct validity	Internal consistency	Test-retest	
ICIQ-OABqol (ICIQ Overactive Bladder questionnaire)[188,189]	A	33 item tool comprehensive tool for overactive bladder. Derived from OAB-q [188]	Men and women	Focus groups, clinical opinion and lit review	✓	✓	✓	✓
ICIQ-Nqol(ICIQ Nocturia Quality of Life Questionnaire)[149]	A+	13 item tool to assess impact of nocturia on quality of life. Previously known as N-QoL [149]	Men and women	4 Focus groups of 7-8 men, clinical opinion and literature review. Only validated in men.	✓	✓	✓	✓
ICSQol[190]	A	8-item tool to assess the impact of LUTS on quality of life in men	Men, LUTS impact on quality of life.	Confirmed that patients understood questionnaires.	✓	✓ (but low)	✓	
ICIQ-LUTSqol (Lower Urinary Tract Symptoms – quality of life)[191–194]	A+	21-item tool, impact of LUTS, associated bother and HRQL. Previously known as King's Health Questionnaire (KHQ) [191]	Men and women	Cognitive debriefing well documented in US population [192]	✓	✓	✓	✓
LIS (The Leicester Impact Scale)[94]	A	21-item tool, impact of urgency, frequency, nocturia and incontinence.	Men and women	Items developed using literature review and discussion with clinicians	✓	✓	✓	✓

3.4 Evaluation of the psychometric properties of selected PROs

The literature search confirmed that that there are no existing, fully validated measures for UAB patient-centered evaluation. The UAB-q provided by the Underactive Bladder Foundation supplies no supportive evidence, published or otherwise, of psychometric properties. However, a number of well validated PROs for the patient-centered evaluation of LUTS and HRQL exist which are likely to capture some of the symptoms experienced by patients with UAB. The current review provides an overview of the extent of published psychometric validation for the selected instruments related to the assessment of LUTS. This also draws attention to where published evidence may fall short of the highest standards of PRO development practice. In addition, the extent to which their published properties are adequate for the standards required by the FDA: 'Guidance for industry for PRO measures: use in medical product development to support labeling claims' [99].

All of the evaluated PROs except the ISS and UAB-q had associated publications which included the method of initial development of items for the questionnaire. The use of a literature review, existing questionnaires, a panel of experts and discussion or interviews with patients were mentioned as the sources by which items were devised, and to ensure the items included were of relevance to the context of use. Current recommendations for PRO development emphasise the particular requirement for the patient perspective through in-depth interviews early on in the questionnaire development process [9,99,195]. Despite their A or A+ grading status, this may be highlighted as a limiting factor in many of the included questionnaires, as for many of the included questionnaires a literature review and consultation with a panel of experts were the sole sources. The result is that issues of importance to patients may be missed, particularly condition indicators relating to quality of life. Although some of the PROs mentioned patient involvement in item development (ICIQ-MLUTS, ICIQ-FLUTS, ICIQ-OABqol, ICIQ-LUTSqol), only the recently developed Lower Urinary Tract Symptom tool (LUTS) [9,178] included the extensive qualitative documentation of patient input through the analysis and reporting of in-depth patient interviews in-line with current regulatory requirements. However, it may be that extensive interviews were completed but the source evidence is no longer available, or published in the public domain (the case for the ICIQ-MLUTS and ICIQ-FLUTS). Nevertheless it remains that the necessary detailed documentation of qualitative research, through in depth-focus groups, or interviews with patients and subsequent applied qualitative methodology is essential in the development of any new PRO measure [195].

Direct patient input is also crucial in order to inform the appropriate patient-centered terminology when developing the items of the questionnaire. In addition, the involvement of patients in the process of cognitive debriefing interviews used to refine the draft instrument are an essential component of demonstrating content validity [196]. Current regulatory requirements require rigorous documentation of patient understanding of all concepts, response options, recall period and overall readability [99]. The ICIQ-LUTSqol (King's Health Questionnaire) demonstrated this aspect of content validity by a well evidenced retrospective cognitive debriefing study. In-depth interviews with 24 patients in the US provided feedback on the instructions, items and response items [192]. The use of subject feedback to confirm patient understanding was also described in the OAB-SS and ICSQol publications [182,197]. The DAN-PSS-1 paper was regarded to have demonstrated face validity as a result of a preliminary trial with patients [184]. The LUTSS publication included some description of revisions made following the feedback supplied by 30 patients who completed the first version of the questionnaire [177]. However, in general, the current review found limited evidence of modifications made as a result of feedback from patient interviews. Evidence from patient cognitive debriefing interviews (e.g. a list of actions taken as a result of patient feedback, concepts elicited by an item) can help confirm content validity by providing evidence that a concept is adequately captured and understood [99].

Aspects of construct validity were demonstrated in some capacity for all of the PROs evaluated. This was achieved by the consistency of responses with a measure which is known to assess the concept of interest in accordance with known theory. For example, the ICIQ-MLUTS was able to detect the expected increase in severity of urinary symptoms with age in a community sample [185]. Several of the measures demonstrated the degree to which the scale did not correlate with a measure designed to assess dissimilar constructs (discriminant validity), or correlates with a measure that should be related (convergent validity). For example, the DAN-PSS-1 showed good correlation with the extensively used Madsen-Iversen symptom score for prostatic obstruction. The Leicester Impact Scale (LIS) was compared with both the Hospital Anxiety and Depression Scale and the Negative Affect Subscale of the Bradburn Affect Balance Scale and correlations were found to be statistically significant. It is important that a PRO tool is able to clearly discriminate between patients from control subjects. This meaningful detection of symptoms was confirmed in the DAN-PSS-1 by comparison of a group of men with no urinary symptoms to a group diagnosed with BPH. The KHQ (now the ICIQ-LUTSqol) was considered a sufficiently validated tool for this to be used as a reference measure for evidence of criterion validity. The B-SAQ was compared directly with

the symptom severity scale of the KHQ, returning high Pearson's correlation values [180]. Evidence of construct validity is evaluated by the FDA when considering a PRO measure, including criterion validity when appropriate.

Elements of reliability (the ability for the instrument to be stable, consistent and reproducible over time) were assessed in all the available publications of the reviewed PRO measures. This property is vitally important for any PRO, particularly for one which aims to measure a long-term chronic condition over time, either in research or clinical practice. From the FDA perspective these properties are essential for instruments intended for use as an outcome measure in clinical trials, in order to provide reliable and credible estimates of patient-reported treatment effect [99]. Internal consistency (the relationship between items) should be high if a measure is able to be scored as a single coherent group. This may be assessed using Cronbach's α statistic, for example high level of internal consistency was observed between the 7 OAB-SS items (Cronbach's α = 0.83) [182]. However, the ICSQol showed poor internal consistency (Cronbach's α = 0.59) [197] demonstrating that the items, although related, should be considered separately. This would make the PRO unlikely to be suitable for clinical trial purposes due to the general desire for score comparisons. Reproducibility, the ability of a questionnaire to produce stable results when the condition of a patient has not changed, may be assessed by a test-retest study. For example, the ICS*male* (ICIQ-MLUTS) was tested on a sub-group of 40 patients who were asked to complete another questionnaire at home within 2 weeks of completion in clinic [185]. The test-retest reliability for the ICIQ-FLUTS was also good; 78% answered identically and 97% within one response category when administered with a two week interval [150]. This time period between administrations should be chosen carefully to reflect the true condition in stable patients and avoid natural variability (for example, anticipated effects of the menstrual cycle) but also in order to minimise memory effects [99].

The responsiveness of the questionnaires was evaluated in 9 of the 15 PROs included in Table 5 and Table 6. This is the ability of an instrument's sensitivity to detect change, and is tested in response to an intervention in which the patient's state has altered with respect to the concept of interest. For example, the responsiveness of the DAN-PSS-1 was evaluated in 29 patients following a prostatectomy and found a median reduction in score of 80%, demonstrating good sensitivity to change. The ICIQ-FLUTS (BFLUTS) was evaluated alongside a randomized controlled trial comparing tension-free vaginal tape with colposuspension in 344 women with stress urinary incontinence [163]. Details were documented of the modifications made to the questionnaire using responsiveness data evidence, including the use of factor analysis to understand the clustering of items to develop a suitable scoring system. However,

the specificity of the responsiveness data to a population of women consenting to a surgical trial for stress incontinence is stated as a limitation of the study. Further testing may be required to determine the responsiveness of the questionnaire in less severe incontinence groups, including those with other LUTS. Indeed, the consequences of an inability to detect change may be to reduce any detection of treatment effect. The extent to which the instrument's ability to detect change in different patient subgroups (e.g. gender, ethnicity, age) should be known, in order that any differences may be taken into account [99].

3.5 Content of selected PROs

The full item wording was obtained for each questionnaire of interest for analysis of the content. Items were categorised to broader symptom concepts such as 'daytime frequency', 'intermittency' or 'hesitancy'. The same process was applied to the questionnaires which measured the impact on HRQL. The aim was to gain some insight into the symptoms, signs and HRQL measurement concepts and to allow comparison between the different PRO tools. An awareness of the content of existing PROs also serves a function of allowing the identification of any new concepts or symptoms which emerge as a result of the development of a more specific UAB PRO. The symptoms and signs for the symptom PROs could be categorised into 14 broad concepts shown in Table 7. Although details of the items which relate to these concepts are not included in the tables, the items and their response options were related to the associated symptom frequency, severity and context. For example, 'incontinence' groups together all items which explore the frequency, type (e.g. stress, urge, enuresis, during physical activity) or severity of the leakage of urine. 'Urgency' groups together all items which explore the circumstance, frequency and severity in which the patient may experience urgency. Similarly, the PROs measuring the impact of symptoms on quality of life had items which could be categorised into 11 main concepts, shown in Table 8. For example, 'journey planning' covers all questions which explored the necessity to 'carefully plan your journey' around the location of toilets, or ability to travel. 'Activities or hobbies' includes any item which aims to identify any impact on hobbies, exercise and sport or other daily activities (e.g. housework, shopping).

The recently developed LUTS tool [9,178] had the most comprehensive coverage of the symptoms with items related to 11 out of the 14 concepts in Table 7. The ICIQ-LUTSqol covered all 11 of the HRQL concepts categorized in Table 8. The questionnaires which are designed to assess overactive bladder (e.g. OAB-SS, OAB-V8, ICIQ-OABqol) have a particular focus on questions relating to urgency and incontinence, associated with the condition. Questionnaires for

patients presenting with LUTS or BPH (DAN-PSS-1, ICIQ-MLUTS, LUTS) have additional items associated with flow, hesitancy and straining. The I-PSS does not include some LUTS which have considerable burden, most notably urinary incontinence [198]. The single questionnaire which is specific to UAB (UAB-q) has items relating to the symptoms of daytime frequency, nocturia, straining, sensation of incomplete emptying and retention, so evaluates no new concepts specific to the target population.

Table 7. Content of selected PROs for the assessment of symptoms.

	Daytime Frequency	Urgency	Incontinence	Nocturia	Hesitancy	Slow stream	Intermittency	Straining	Post-micturition dribble	Incomplete emptying	Retention	Pain	Continence aids	Splitting or spraying
B-SAQ	✓	✓	✓	✓	×	×	×	×	×	×	×	×	×	×
OAB-SS	✓	✓	✓	✓	×	×	×	×	×	×	×	×	×	×
DAN-PSS-1	✓	✓	✓	✓	✓	✓	×	✓	✓	✓	×	✓	×	×
ICIQ-MLUTS	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	×	×	×	×
ICIQ-FLUTS	✓	✓	✓	✓	✓	×	✓	✓	×	×	×	✓	×	×
ISS	✓	✓	✓	✓	×	✓	×	✓	×	✓	×	×	✓	×
OAB-V8	✓	✓	✓	✓	×	×	×	×	×	×	×	×	×	×
LUTSS	✓	✓	✓	✓	Information not available									
LUTS	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	×	×	×	✓
I-PSS	✓	✓	×	✓	×	✓	✓	✓	×	✓	×	×	×	×
UAB-q	✓	×	×	✓	×	×	×	✓	×	✓	✓	×	×	×

Table 8. Content of selected PROs for the assessment of associated HRQL.

	Journey planning	Social life	Family life	Activities or hobbies	Sexual activities	Tiredness	Emotional impact	Hygiene	Fluid intake	Continence aids	Overall QoL
ICIQ-OABqol	✓	✓	✓	✓	×	✓	✓	✓	×	×	✓
ICIQ-Nqol	×	×	✓	×	×	✓	✓	×	✓	×	✓
ICSqol	×	×	×	×	×	×	✓	×	✓	✓	✓
ICIQ-LUTSqoL	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
LIS	✓	✓	✓	✓	×	×	×	×	×	×	×

3.6 Conclusion

The evaluation of the extent of the published psychometric evidence of the selected PROs highlighted a plethora of questionnaires, some of which have been robustly validated. However, evidence of patient involvement in the initial development and refinement of items, aspects of reliability, and the testing of instrument responsiveness in multiple patient sub-groups was lacking (although it is acknowledges that the latter is often the subject of ongoing development). The continual development of PRO measure good practice guidelines ensure PROs accurately capture patient-centered information for clinical practice and research, as well as enabling the rigorous use of tools as outcome measures in clinical trial settings. The new LUTS questionnaire [9,178] shows promise as the first questionnaire for the assessment of LUTS to be developed with published evidence of content validity according to FDA regulatory standards [99].

The current review of the existing literature has confirmed that no fully validated PRO measures exist which have been specifically developed for the assessment of UAB. The UAB-q is the only available tool which specifically aims to assess the symptoms of UAB but there is no published evidence of its psychometric properties. A number of PROs which evaluate LUTS and associated HRQL to a high degree of evidenced validity were identified. The content of these measures was evaluated in order to gain insight into the symptoms and concepts covered by existing PROs in this area. Although the LUTS tool, ICIQ-MLUTS and ICIQ-LUTSqol are likely to capture some of the UAB symptoms and impact, these PROs were developed for a specific context of use. The inclusion of items, including the wording and content of the PRO measure, should be based on the patient-centered input of the specific population of interest. The understanding of the patient experience of underactive bladder is so little understood, this supports the qualitative exploration of this during the development phase of a PRO for UAB. In addition, evidence of its utility in the target population is required in order for a PRO to be considered a credible measure for research and clinical practice [112]. From a clinical trial perspective, in order for a PRO measure to be used to support a labeling claim, it must be shown to reliably measure the concept in the patient population in which the clinical trial is to be conducted [99]. This is also important in clinical practice where treatment decisions may rely on PRO evidence to reflect the patients' perspective. In conclusion, the lack of an existing tool validated in the UAB target population to current regulatory standards supports the development of a novel PRO instrument for the assessment of underactive bladder.

Chapter 4 Concept elicitation and development of the initial draft ICIQ-UAB

4.1 Introduction

The literature review in chapter 1 established that the clinical symptoms of DU are not well defined, although it is recognised that there is a significant clinical overlap with many of the symptoms experienced by patients with BOO and overactive bladder (OAB) [21]. These symptoms include high urination frequency, nocturia, urgency, and incontinence [14,25]. Other symptoms suggested to be associated with DU include hesitancy, a sensation of incomplete emptying, straining, a weak stream and persistent urinary tract infections (UTIs) [25].

As described in chapter 3, no PRO instruments currently exist, to assess UAB symptoms and impacts, that would meet the standards of validation described by the FDA's guidance for Industry [99]. The International Consultation on Incontinence Questionnaire (ICIQ) modules offer a range of psychometrically robust instruments for the self-assessment of lower pelvic dysfunction including LUTS [199]. Due to the recognised need for a new specific PRO measure for UAB, the Bristol Urological Institute (BUI) initiated the development of a new ICIQ module, the ICIQ-UAB, with the aim of capturing the patient reported symptoms and impact of the condition UAB, for eventual use as an outcome measure in future clinical trials and in clinical practice.

This chapter describes the consultation with an expert clinical panel of experts and the exploration of the patient experience of UAB by semi-structured interviews with patients. In addition to the literature and PRO instrument review described in chapters 1 and 3, this informs item generation for the developmental instrument and supports content validity [99,114]. As data is collected, the developmental conceptual framework and quality of life components of the disease model are revised. The chapter then describes the rationale for the choice of items, response options, format, wording and recall period and presents the first version of the ICIQ-UAB for ongoing development.

4.2 Clinical review panel

A nine-membered clinical panel consisting of one nurse, two clinical scientists and six urologists (Appendix 1) were recruited for their experience in urodynamic assessment from different sites across 4 NHS trusts over the UK. All those were invited to be on the panel agreed to be involved. Each member was consulted on hypothesised concepts of relevance to UAB. A cross-section of health professionals were chosen to represent different perspectives on the condition:

Urology Clinical Nurse: Specifically, a nurse providing assistance with catheter use, as they will have significant contact with those patients suffering with retention.

Clinical Engineer: This role has a functional knowledge of urodynamic assessment and diagnosis.

Urology Clinician & Senior Urology Clinician: This role benefits from an in-depth knowledge of urological conditions, developed over a number of years.

A pragmatic approach was used to collect the opinions from the clinical experts. Initially, a list of hypothesized concepts of relevance derived from the literature review were proposed by email, whereupon they provided feedback at the level of 'essential', 'desirable' or 'not required'. This feedback was used to generate the initial interview schedule and conceptual framework. Additional comments or suggestions for items were also recorded and incorporated into the development of the initial conceptual framework. Members of the clinical panel were then involved in the generation of the individual items before the cognitive interviewing.

Table 9 shows that most clinical panel members considered the number of urinations after waking (before sleep), the number of urinations after sleep (before waking), the strength of flow rate, the use of straining to start and/or to continue voiding, intermittency, post-micturition dribble, the sensation of incomplete emptying, the awareness of leakage, and the sensation of urgency as 'essential' or 'desirable' to include in a PRO instrument for the assessment of UAB. Green shading indicates when there was consensus and amber when there was no clear consensus.

Table 9. Clinical panel rating of proposed symptoms and impacts to include in the ICIQ-UAB.

Item	Essential	Desirable	Not required	No Entry
Symptom Component-				
Number of urinations after waking, before sleep	4	2	1	1
The clustering of urinations, for example in the morning	0	2	5	1
Number of urinations after sleep, before waking	3	3	0	2
Strength of flow rate	6	1	1	0
The use of straining to start voiding	6	1	1	0
The use of straining to continue voiding	6	1	1	0
Intermittency	6	1	0	1
Hesitancy	6	0	1	1
Post-micturition dribble	3	3	1	1
Sensation of incomplete emptying	5	2	1	0
Is the individual aware of leakage, when it occurs?	0	4	3	1
Does the individual experience the sensation of urgency?	2	2	3	1
How long after the first desire to void, before feelings of urgency, measured in...?	1	0	6	1
Impact components				
Sleep/Rest	4	4	0	0
Physical activities	3	2	3	0
Social impact	3	5	0	0
Psychological wellbeing	2	5	1	0
Planning & task management	1	6	1	0

4.3 Preliminary conceptual framework

The expert clinical review panel, literature and instrument review (chapters 1 and 3), informed the development of a preliminary conceptual framework (Figure 3). This defines the concepts measured by the instrument in a diagram which presents a description of the relationships between items, domains (sub concepts), in the PRO instrument [99]. The conceptual framework informed the interview guide for the initial qualitative concept elicitation interviews.

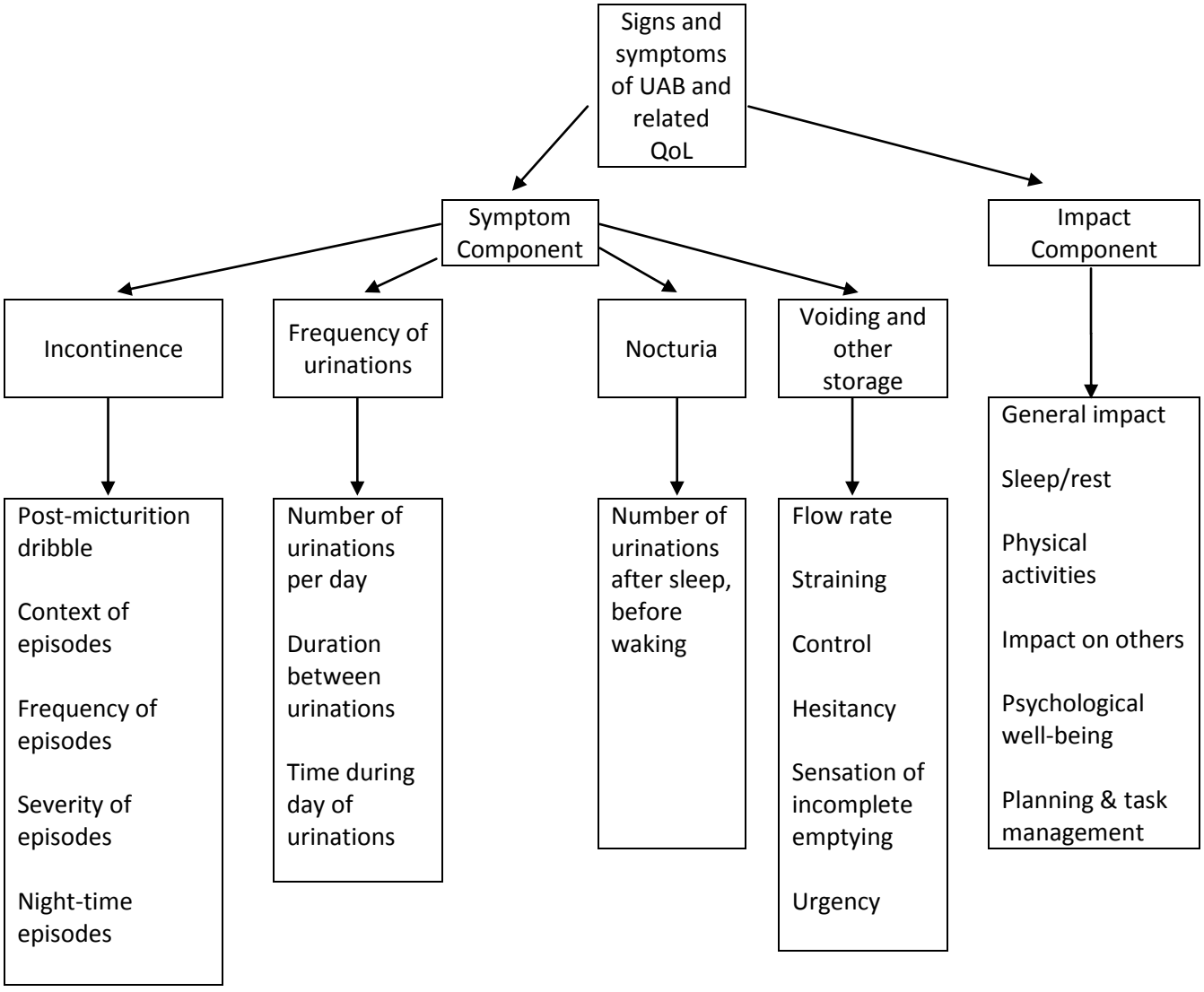


Figure 3. Preliminary conceptual framework.

4.4 Concept elicitation

4.4.1 Aims and objectives

Qualitative interviews with patients who had been diagnosed with DU were used to elicit the patient experience of associated urological symptoms and their impact on day-to-day life. Individual interviews were chosen over focus groups as they are better suited to topics which are personal or sensitive in nature or when a discussion requires particularly detailed experiential information on a topic which is little understood [112]. Specific objectives included:

- The documentation of the patient reported signs, symptoms, and impacts in patients with urodynamically confirmed DU.
- The exploration of patients' words to describe the severity, frequency, duration and impact of patient reported signs or symptoms.

4.4.2 Sample size

It was anticipated that 40 interviews would be required in order to achieve sufficient concept saturation. This sample estimate was based on previous experience and the literature [9,121,124,125]. This is a relatively large sample for a qualitative interview study, but was considered appropriate to allow the exploration of a relatively heterogeneous sample. As patients diagnosed with DU often present with other co-existing urological conditions, those with co-existing DO, SUI, urodynamic stress incontinence (USI), bladder outlet obstruction in the equivocal range (BOO-E), and BOO, were also interviewed.

4.4.3 Site selection and recruitment

Patients were recruited at the Bristol Urological Institute (BUI), Southmead Hospital, Bristol. This is a tertiary referral centre with increased access to patients with a wide range of symptoms. The BUI also has a unique database that contains detailed urodynamic data for all patients who have attended the site over the last 30 years, giving a large pool of potential patients.

Patients were identified based on their historical urodynamic records, accessed through the Clinical Information System Suite at the North Bristol NHS Trust (NBT). The Urodynamic clinic's Daybook was used on an ongoing basis in order to recruit patients who recently underwent urodynamic assessments.

4.4.4 Inclusion and exclusion criteria

Identification of potential patients for recruitment to the study was according to the following inclusion criteria:

- Male and female patients, 18 years old or older
- Registered at Southmead Hospital, Bristol
- A confirmed urodynamic diagnosis of DU within the last eighteen months
- Able to speak, read, and write in English
- Willing and able to participate in a face-to-face interview session.

The following were excluded:

- Patients with disorders interfering with the study conduct (such as some psychiatric disorders, malignant disease)

The pressure flow study (PFS) parameters used to identify potential patients were the bladder contractility index (BCI) and bladder outlet obstruction Index (BOOI) for males, and detrusor pressure at maximum flow ($p_{det}Q_{max}$), and maximum flow rate (Q_{max}) for females. Patients were categorised in the diagnostic groups to facilitate the analyses of the data, as detailed in Table 10.

Table 10. Diagnostic group inclusion criteria.

DU (only) • Patients with a confirmed diagnosis of DU on urodynamics and a diagnosis of no other co-existing urological conditions Males: BCI <100 BOOI <20 Females: $p_{det}Q_{max}$ <20 cmH_20 Q_{max} <15ml/s
DU + mild DO • DU + co-existing mild DO (based on investigator urodynamic assessment and patient records)
DU + Mild SUI/USI • DU + co-existing mild USI/SUI (based on investigator urodynamic assessment and patient records)
DU + BOO-Equivocal (males only) • BCI <100, BOOI ≥20–<40
DU + BOO (males only) • BCI <100, BOOI ≥40
BOO (males only) • BCI >100, BOOI ≥40

Abbreviations: Detrusor pressure at maximum flow ($p_{det}Q_{max}$), maximum flow rate (Q_{max}), bladder contractility index (BCI) calculated by BCI=$p_{det}Q_{max}$+5Q_{max}, Bladder Outlet Obstruction Index (BOOI) calculated by BOOI=$p_{det}Q_{max}$-2Q_{max}.

4.4.5 Concept elicitation study process

Patients who met the inclusion criteria were sent a postal invitation, the patient information sheet (Appendix 2), a stamped addressed envelope, and a form to express their interest to take part. Patients were given the option of having their interview at their home, in the hospital, or over the telephone. In each setting, the interview was conducted in a room which minimised the risk of interruption and maintained the privacy of the patient. For telephone interviews, recorded verbal consent was taken and the interview recorded whilst on speakerphone in a private room.

4.4.6 Interviewers and quality assurance process

The first 44 interviews were conducted by a study researcher, previously assigned to the project. The author then assumed full responsibility for the remaining interviews and analysis. Transcripts were coded on an ongoing basis to inform study progression. Regular meetings between the project supervisor and the interviewers were held in order to discuss the interview process, content and technique. Upon completion of all interviews, all transcripts were independently analysed by the author.

4.4.7 Interview guide

The interview guide passed through four iterations, comprising of minor additions to allow further exploration of emerging concepts in subsequent interviews. The final version is included in Appendix 3. Open ended questions were favoured such as 'Which of your symptoms do you find most bothersome?', instead of specific probes, to elicit more spontaneous responses.

4.4.8 Ethics and informed consent

The study was conducted with favourable opinion from the Southmead Research Ethics Committee (REC), now the Bristol South REC: REC reference 087/99. For face-to-face interviews (n=20) the patients read and signed an informed consent form (Appendix 4). For telephone interviews (n=27) the patients were read a telephone informed consent procedure,

whereupon they provided a verbal response which was recorded at the beginning of the interview.

4.4.9 Data management

All interviews were audio recorded and transcribed verbatim. Transcripts were checked for accuracy using referral to the original audio recording and analysed using NVivo v10, a qualitative data software package. Files were stored on the NBT secure server and all patient data was anonymised using a unique study identifier. A hard copy of contact details and unique identifier, along with a case report form were kept in a locked office on NBT premises.

4.5 Concept elicitation results

4.5.1 Sample demographic and urodynamic characteristics

In total, 44 patients with DU were interviewed; 32 of which were male and 15 female. All patients were White British. The age range was 27 to 88 years, with a mean of 64 years. The patient characteristics summarised by diagnostic group are given in Table 11. Twelve patients provided information regarding educational background. These patients came from a variety of educational backgrounds (ranging from left school at 16 or younger, college or university educated) and occupations (manual, service and professional). In addition to the 44 patients with DU, three patients with BOO were interviewed in order to explore possible differences in symptoms experienced by patients with DU and patients with BOO. The number of invited participants was not recorded at this stage (by the previous researcher who conducted the concept elicitation interviews). However, reasons given for any non-participation were primarily that the patient was non-contactable by telephone or post. Although the non-recruitment of any 'true decliners' is acknowledged as a limitation, the intention for this study was to elicit the patient experience, so any decliners were considered unlikely to significantly affect the results due to the planned continuation of recruitment to theoretical data saturation.

Table 11. Summary of demographic and clinical characteristics.

Clinical or demographic characteristic	Total sample (n=44)	DU	DU + co-existing urological conditions
DU DU + DO DU + SUI DU + BOO-E DU + BOO	19 8 7 5 5	19	25
Mean age and range (years)	64 (27-88)	59 (27-88)	68 (38-87)
Gender, male *n* (%)	29 (66)	12 (63)	17 (68)
Intermittent self-catheterisation *n* (%) (Historical or current)	23 (52)	10 (53)	13 (52)
PVR >30ml* *n* (%)	34 (77)	14 (74)	20 (80)
PVR >30ml* (ml) (median and IQR)	199 (100-492)	335 (119-492)	170 (100-360)
BCI (median and IQR)	62 (49-79)	62 (48-82)	62 (50-77)
BOOI (median and IQR)**	18 (8-28)	15 (6-18)	25 (9-41)
$p_{det}Q_{max}$ (cmH20) (median and IQR)	25 (12-35)	24 (12-29)	26 (12-36)
Q_{max} (ml/sec) (median and IQR)	8 (6-10)	8 (6-11)	6 (5-9)

*In the absence of any evidence base for the lower limit of a 'significant' PVR we chose >30mls.
**Males only
Abbreviations: Detrusor pressure at maximum flow ($p_{det}Q_{max}$), maximum flow rate (Q_{max}), bladder contractility index (BCI) calculated by BCI=$p_{det}Q_{max}$+5Q_{max}, Bladder Outlet Obstruction Index (BOOI) calculated by BOOI=$p_{det}Q_{max}$-2Q_{max}, Post Void Residual (PVR).

4.5.2 Qualitative coding

Following familiarisation with the data, the transcripts were inductively coded [134] to reflect the content of the interviews. This involved identifying sections or phrases of text and categorising quotes into an ongoing coding framework. Data collection and analysis continued concurrently, using a constant comparison approach informed by grounded theory [133]. Any concept or code that was considered spontaneously reported in the interview (without prompting by the interviewer) was coded separately. Towards the end of data analysis and during reporting, concepts were further defined by known urological symptoms (e.g. 'incontinence', 'nocturia', 'straining) [200,201] which organised the data within the relevant

theoretical context according to the principles of thematic analysis [126]. It is acknowledged that the reliability of the findings is dependent on the continual reflexive appraisal by the researcher and referral to the original data during the analytical process [127]. The themes and sub-themes were backed up by multiple quotations from different respondents wherever possible, in order to provide alternative perspectives of the same contextual reality and to lend credibility to the conclusions.

4.5.3 Saturation of concept

The transcripts of patients with DU without co-existing urological conditions were analysed first, in the order in which they had their interviews. This was so the developing codes were defined primarily by the diagnostic group of primary interest. Data saturation was achieved by the first 19 interviews with DU patients (Table 12). The additional groups with co-existing urological conditions were then analysed to achieve full exploration among those who represented potential respondents. There continued to be a small number of new codes from these groups. However, upon inspection these were considered to be sub-concepts more likely to be related to the co-existing conditions. Additional interviews were unlikely to result in further concepts relevant to DU, which fulfils the criteria of data saturation in PRO instrument development [120,121].

Table 12. The saturation of concept by order of analysis and diagnostic group.

Order of analysis	Patient number	Diagnostic group	Number of new codes
1	P2	DU	22
2	P11	DU	24
3	P12	DU	20
4	P13	DU	14
5	P14	DU	5
6	P15	DU	2
7	P16	DU	9
8	P17	DU	7
9	P18	DU	2
10	P20	DU	4
11	P21	DU	5
12	P22	DU	8
13	P23	DU	0
14	P32	DU	2
15	P34	DU	1
16	P35	DU	2
17	P36	DU	1
18	P37	DU	0
19	P39	DU	0

Order of analysis	Patient number	Diagnostic group	Number of new codes
20	P1	DU & DO	2
21	P3	DU & DO	1
22	P6	DU & DO	4
23	P10	DU & DO	2
24	P19	DU & DO	1
25	P30	DU & DO	0
26	P33	DU & DO	2
27	P38	DU & DO	0
28	P4	DU & SUI	0
29	P5	DU & SUI	0
30	P7	DU & SUI	1
31	P8	DU & SUI	2
32	P9	DU & SUI	3
33	P26	DU & SUI	0
34	P29	DU & SUI	2
35	P24	DU & BOO-E	1
36	P25	DU & BOO-E	0
37	P27	DU & BOO-E	0
38	P28	DU & BOO-E	1
39	P31	DU & BOO-E	0
40	P40	DU & BOO	1
41	P41	DU & BOO	0
42	P42	DU & BOO	0
43	P43	DU & BOO	1
44	P44	DU & BOO	0
45	P45	BOO	0
46	P46	BOO	1
47	P47	BOO	0

4.5.4 Symptoms

Twenty distinct lower urinary tract symptoms, signs or experiences were described by the patients. For each, a qualitative description of the reported patient experience with supportive patient quotes from the interviews is given. If an ICS urological definition exists for the described symptom then this is included within the description. The number of patients who experienced each symptom is given separately for the DU sample, and for those with co-existing urological conditions.

Daytime frequency of urination

The definition provided by the ICS for increased daytime frequency of urination is '*the complaint by the patient who considers that he/she voids too often by day'* [200]. Most patients from the DU sample (13/19) and slightly less than half of those with co-existing urological conditions (12/25) reported high urinary frequency, often spontaneously, and as one of their most bothersome symptoms.

> P22: "it's the keep having to go to the toilet all the time"
> P34: "Just sort of constantly going back and to and from the toilet sort of it's just sort of a pain really"

However, a small number in the DU sample (n=3/19) described either 'normal' or low frequency of urinations (once or twice a day).

> P11: "No I can go quite long periods without going."

Although for most patients there was no difference in time of day, some patients (n=4/19 in the DU sample) described having to urinate more often in the morning or many times in short succession immediately after rising from bed.

> P16: "When I wake up I have a cup of coffee and then go again and then again and then maybe again."

Hesitancy

The definition of hesitancy provided by the ICS is 'an individual describes difficulty in initiating micturition resulting in a delay in the onset of voiding after the individual is ready to pass urine' [200]. This was a common symptom reported spontaneously by most patients of either gender in the DU sample (13/19) and amongst the other groups with co-existing urological conditions (11/25). Patients described the process of 'having to wait' or 'concentrate' before

the flow of urine would start. The process was a combination of relaxing and 'getting into the right frame of mind' in order to facilitate the onset of urination.

> P11: "You stand there you can't go. You take a while to go."
> P11: "It seems to travel a long distance before it comes out and I stand there sometimes and I think I've got to concentrate for it to come out."

The delay for most patients was usually a few seconds but for some individuals there could be a wait of up to twenty minutes.

> Researcher: "How long do you have to wait sometimes?"
> P35: "Sometimes twenty minutes."

Urinary incontinence

The ICS definition of urinary incontinence is *'the complaint of any involuntary leakage of urine'* [200]. This symptom occurred in a few of the DU patients (5/19) but was not often mentioned spontaneously in this group. Circumstances of incontinence in the DU group was generally when they felt they were 'cut short' or that they 'did not quite make it in time' to the toilet. Leakage described in the DU sample was otherwise reported as occurring more occasionally and in 'small amounts'.

> P35: "Sometimes I do and sometimes it leaks and I've had a few leakages that's my fault though for trying to hold it in too long."

Episodes of incontinence were often more severe and frequently occurring in the other groups with co-existing urological conditions (15/25), and particularly in patients with DO and SUI. The impact and level of 'bother' described by patients experiencing incontinence was often very high.

> P7: "I don't feel like I need to go it just literally it's involuntary if you know what I mean it just kind of happens. I've got no control over it whatsoever."

Nocturia and/or nocturnal voids

Nocturia, *the complaint that the individual has to wake at night one or more times to void*' [200] was a common symptom in the DU only group (15/19) and the group with DU and co-existing urological conditions (19/25). It was frequently talked about spontaneously as one of their most bothersome symptoms.

> P20: "Yes I used to wake up, well I still do wake up a couple of times at night to go to the toilet and I just presume this is perfectly normal."
> P21: "The worst thing was getting up at night um constantly up and down up and down up and down."

The number of nocturia episodes was variable between patients, one DU patient described having to get up to urinate as many as nine times a night, but most described getting up out of bed between one and four times a night to pass urine. Some would know when and how many 'almost to the minute', however for others the number of times per night showed day to day variability.

> P22: "One night it can be twice and another night it can be five times you know, you just don't know."

Sensation of incomplete emptying

Patients described the feeling that urine remained in the bladder soon after urinating, often resulting in having to return to the bathroom again within a short period of time. Although not often mentioned spontaneously, this symptom was often present in all groups (8/19 in the DU group and 11/25 in those with co-existing urological conditions.

> P34: "Sometimes it's just I know instantly that in a couple of well, 10 minutes, I'm gonna need to go again."

Some patients described their tendency to wait in the bathroom or continue sitting down on the toilet for a few minutes after urinating in order to empty their bladder as much as possible.

> P3: "You've got to sit there a bit longer, like, because you know you're going to do a bit more."

Others suspected they were not emptying their bladders properly each time they urinated, due to voiding only small volumes per urination.

> P11: "When I do pass it's not a lot. It's not a tremendous amount."

Slow stream

The strength of flow is clarified by the ICS as *'his or her perception of reduced urine flow, usually compared to previous performance or in comparison to others'* [200]. This was a very common symptom with both male and female patients describing, often spontaneously, a 'slow' or 'weak' flow with an associated prolonged urination time (13/19 in the DU group and 15/25 in those with co-existing urological conditions).

> P39: "I can pass urine with a reasonable stream at sometimes and sometimes it's sort of just limps out and I can never know which one it's going to be to be honest"

Some male patients described the flow as coming out as a 'vertical drop' or just going 'straight' down. The flow rate could also be variable during the process of passing urine and dependent on the fullness of the bladder or strength of the urge to go.

> P22: "Sometimes it start of it slow and then it suddenly becomes faster and other times it will come out fine and then slows right down."

Reduced sensation of bladder fullness

A few of the DU patients (2/19 and 3/25 in the groups with co-existing urological conditions) would feel empty after urinating but nevertheless could not be sure that they had fully emptied the bladder. There was a perceived lack of sensation of the fullness of the bladder which caused them to doubt whether they had emptied properly, despite not having any sensation of a residual.

> P35: "Yeah it does feel like I have emptied it yeah but I don't get- it's difficult to tell because I don't get the sensation of full or not full"

Acute retention

Some patients experienced retention: *'patient is unable to void without catheterisation'* [200] (4/19 in the DU only group and 2/25 in the group with co-existing urological conditions). This was the clinically termed 'acute' retention pertaining to the hospitalisation of a patient who is unable to urinate at all, and the insertion of a catheter is required to drain the bladder. In these cases, it was often volunteered as the event for which they first sought medical assistance for their urological symptoms.

> P35: "Can't go to the toilet on full bladder can't wee without running the taps trying everything sitting on the toilet showers things like that to try and help but nothing would come out."

Temporarily unable to pass urine

For some patients (4/19 in the DU group), there were infrequent circumstances in which they were unable to voluntarily pass urine, such as whilst standing at a urinal, during the night, or when asked to pass urine by a medical professional. These patients reported that they would usually return to the bathroom a few minutes later to try again.

> P2: "I mean sometimes, you know, you go standing up, no problem at all, everything's fine. Other times, I could be in there, stood for half a minute or two, nothing. I think, well, alright, I'll go and do summat and I'll come back a minute later, still standing up and I can do it, I can go, no problem."

Straining

The ICS defines straining as *'The muscular effort used to initiate, maintain or improve the urinary stream'* [200]. A large proportion of the DU group (14/19) and 9/25 in the groups with co-existing urological conditions reported, often spontaneously, the 'pushing', 'straining' or 'squeezing' when starting, during or finishing urination. Straining was generally a behaviour which was a consequence of a number of other symptoms. Patients would strain to start urination if there was a delay (hesitancy), to maintain their flow or in order to restart urinations when it stops (due to an intermittent stream), or to try and 'push' any perceived remaining urine out.

> P12: "There would be very little natural flow and the majority of the flow would be as a result of having to strain."
> P6: "I do try to squeeze every last drop out so I don't have any accidents."

Intermittency

Intermittency is defined as 'when the individual describes urine flow, which stops and starts, on one or more occasions during micturition' by the ICS [200]. Patients described the flow of their urination not being continuous with multiple breaks of a few seconds. This was a common symptom in the DU sample (11/19) and occurred in the groups with other co-existing urological conditions in 6/25 of the interviews.

> P39: "It will take several starts and stops for it to get rid of that particular bladder full"

Post-micturition dribble

The ICS defines dribbling as when an 'individual describes the involuntary loss of urine immediately after he or she has finished passing urine, usually after leaving the toilet in men, or after rising from the toilet in women' [200]. Patients in the DU sample (7/19 and 10/25 in the groups with other co-existing urological conditions) described the leakage of a small amount or 'a few drops' of urine shortly after urinating and having got dressed. This symptom was described as bothersome as patients worried about smell or having a small damp patch on their trousers. Only one female patient mentioned this as a bothersome symptom in the sample.

> P39: "You think you've shaken everything off get back in to bed and there's just one drop or two you know leaks out."

Urgency

This is a symptom which is defined as *'a sudden compelling desire to pass urine which is difficult to defer'* by the ICS [200]. Patients described the sudden onset of a strong desire to urinate which they were unable to ignore and had to 'rush to the toilet'. Patients in every group experienced urgency in accordance with this description (11/19 in the DU group and 11/25 in the groups with other co-existing urological conditions).

> P12: "It's almost coincidental the trigger in your mind; ah I think I need to go to the loo within seconds its saying let me get there quickly"

Lower urinary tract pain

Bladder pain as defined by the ICS *'is felt suprapubically or retropubically, usually increases with bladder filling and may persist after voiding'* [200]. Several of the DU patients (6/19) and 5/25 in the groups with co-existing urological conditions described pain but there was variation in their accounts of the type, source and level of pain experienced. When prompted, patients described the location or origin of the pain as 'across the stomach', or in the 'pelvic' or 'kidney' area. The level of pain described ranged from sharp and 'excruciating' to mild 'discomfort'.

> P36: "Terrible- it's terrible I normally get sharp sharp pains if I have to hold it in I gotta go straight away."
> P18: "It's usually when I've got a full bladder and then it hurts when I've emptied the bladder."

Spraying

A few male patients (4/19 in the DU group and 5/25 in the groups with co-existing urological conditions) reported a lack of control over the direction of their urinary stream. This included spraying or involuntary 'sideways' splitting of the stream. Various strategies were reported to reduce this inconvenience including four patients regularly urinating into a jug and several men choosing to sit down to avoid 'splash'. No female patients reported this as a symptom.

> P35: "I seem to splash a lot when I go I do spray so it doesn't come out in a straight line it sometimes goes a bit sideways"

Associated bowel symptoms

All patients were asked if they had any bowel difficulties and a small number of both male and female patients (3/19 in the DU sample and 5/25 in the groups with co-existing urological conditions) experienced problems in this area. Two female patients had noticed that when they were constipated, their urinary symptoms were worse and drew a direct association.

P13: “I was going to say my bladder is massively affected by my bowel. So the more constipated that I am it brings all of my bladder symptoms out.”

The other patients who experienced bowel problems did not necessarily relate the process of opening their bowels to their urinary difficulties.

4.5.5 Impacts

The impact of symptoms reported by patients on their day to day lives ranged from extremely severe to very little. Several of the patients described how they adapted their lives around their condition so that any impact was minimised and asserted that they had become ‘used to it’ due to the chronic nature of their condition (5/19 in the DU group and 7/25 in the groups with co-existing urological conditions). Nonetheless, patients described different aspects of their lives which could be inconvenienced or made impossible by the ramifications of their symptoms.

Planning life around location of toilets

Several of the symptoms, in particular high frequency of urination and urgency led to their adjustment of their plans around the location of toilets (13/19 in the DU group and 14/25 in the groups with co-existing urological conditions). Patients reported that when leaving the house they needed to plan toilet stops for long journeys, or to know in advance the location of toilets at the place of destination. Others mentioned the considerable impact that their reliance on the location of toilets had on planning holidays, the adaptations to their work life, physical activities, and aspects of their social lives.

P34: “Well I’m just constantly being aware of where my surroundings sort of making sure there’s a toilet nearby if I’m going to a meeting I’ll go before and then sometimes during or straight afterwards”
P12: “I always have to sort of have a plan in my mind as to where we’re going to go where there are toilets that I can use. Whereas I never used to give it a thought.”

Impact of nocturia and/or nocturnal voids

Tiredness, fatigue and having to have naps in the middle of the day as a result of ‘broken sleep’ were commonly reported by those patients who suffered from nocturia. This had a significant effect on quality of life for many of the patients (7/19 in the DU group and 5/25 in the groups with co-existing urological conditions).

P21: “I really would crave a decent night’s sleep”
P13: “My sleep as such is rubbish I’m always fatigued... I rarely feel full of energy or not tired. Yeah feeling tired is one of the biggest kind of um restraints in my life.”

P36: “Then going to bed for a couple of hours this is what I’ve had to do during the day just to compensate getting not enough sleep at night.”

Feelings about self

Some of the patients described the effect of their condition as making them ‘feel old’, less motivated to do things, and the negative effect on self-esteem and confidence (3/19 in the DU group and 11/25 in the groups with co-existing urological conditions).

“P18: Well it makes me feel old”
“P6: Well I suffer a bit with depression for a while and it was making it worse”

However, most of the patients who gave answers to this in the DU group did not consider their condition to affect the way they felt about themselves. It was reported they had adapted their lifestyles and psychological approach to accepting their condition and hence their symptoms had minimal impact.

P4: “I don’t worry. I mean if you let it get you down it will get you down. But I don’t. You know I mean you’ve got to just carry on and think well, there are other people worse than me.”
P7: “I’ve come to terms with it a bit more now.”

Embarrassment

Situations which caused patients embarrassment were mentioned spontaneously as a demonstration of the effect and inconvenience that their urinary symptoms could have on their lives (4/19 in the DU sample and 7/25 in the groups with co-existing urological conditions). Typically this was due to their high frequency of urination or length of time required in the bathroom. Several male patients mentioned their reluctance to use urinals due to hesitancy or poor flow.

P17: “I’m a bit embarrassed about saying excuse me I just need to go to the toilet again you know”
P12: “A young boy came in emptied his bladder and by the time he’d finished and gone I hadn’t even started and I became a bit embarrassed by it by that situation because I was just stood there doing nothing”

Fluid intake

The monitoring of their fluid intake was often offered spontaneously as an adaptive strategy for many of the patients in order to minimise the effect of their urinary symptoms. This could be either the type of drink or volume and was particularly in relation to alcohol or caffeinated drinks which many perceived as exacerbating their symptoms (8/19 in the DU sample and 7/25 in the groups with co-existing urological conditions). This was seen as having an effect on their

ability to partake in certain social situations and was part of the general inconvenience and effect that their urinary symptoms had on their lives.

> P12: "I deliberately avoided taking any fluids, which made me at times feel quite uncomfortable dehydration and I thought this is getting silly and I need to drink more fluids particularly on a day like today you know where you could dehydrate."

Impact on family and friends including sex-life

This was primarily the impact on a partner of having to get up multiple times during the night. However, others intimated that they felt that symptoms put up a barrier with family members as they found it difficult to talk about their symptoms, effect of low mood on relationships, and the pragmatic inconveniences on others of having to be in close proximity to a toilet (5/19 in the DU sample and 5/25 in the groups with co-existing urological conditions).

> P16: Well I think over the years I just got used to it; its fine, but I think my husband because I wake him up getting up and up.

A question about difficulties of a sexual nature was not included in the interview guide. However, most of the patients who talked about this described no effect on their sex life or that it was no longer an important part of their lives. However, a small number of patients (1/19 in the DU only group and 3/25 in the DU with co-existing urological conditions group) talked about their difficulties in this area and the associated impact on their intimate partners. This was of considerable impact for the few that mentioned these issues.

> P29: "Yeah if you love your wife you want to sort of physically show that and I can't now of course."

4.5.6 Specific groups

Catheter use

The majority of patients in the DU group (10/19) and 13/25 in the groups with co-existing urological conditions were currently self-catheterising or had done so in the past. There were varying reports of the impact on daily life of catheter use. Some described it as very inconvenient and uncomfortable whilst for others it enabled them to continue with their lives without being hampered by the effect of their urinary symptoms. Some patients reported self-catheterising sometimes every time they urinated so up to five times a day. Others would only self-catheterise when necessary so only once or twice a week. Four of the DU only group were

catheterising at least once a day but not every time they urinated. One patient, who performed intermittent self-catheterisation, described some relief of her storage symptoms (particularly nocturia).

> P23: "One of the things self-catheterisation did was that I could go for quite a long period without having to go to the loo. But now I'm having to go what I would say is more normally and I'm not aware that I've got any residual or anything like that and I think I'll go maybe maximum twice in the night maximum sometimes not even once"

Urinary tract infections

Many of the patients mentioned that they had experienced urinary tract infections (UTIs) (8/19 in the DU group and 9/25 in the groups with co-existing urological conditions). The impact on lifestyle and level of bother described was very high when suffering with an UTI and was often mentioned spontaneously without prompting. Urinating was extremely uncomfortable and frequent, with often cloudy or 'smelly' urinations, with some patients having a noticeable effect on their cognitive function.

> P23: "What I did get was a certain feeling of feeling unwell and irritable and more or less um unable to cope basically."
> P2: "The other thing about it, as well, is all the infections you get whilst that's going on, because that is a nightmare as well, because it's agony."

Bladder outlet obstruction

The symptoms the three patients with BOO reported overlapped with those of DU patients. This included the main symptoms of high frequency, nocturia, hesitancy, intermittency, nocturia, and slow flow. One patient reported straining towards the end of urination; another, however, found that straining did not help with flow:

> P46: "Well I do um although I've found I've learnt the opposite is the best way to try and relax rather than strain"

Two of the BOO patients mentioned pain around the prostate area or in the penis as particularly bothersome:

> P46: "The one that bothers me the most is the acute pain get in my prostate gland or in that in that sort of area behind my um behind my testicles by my anus it's that sort of area that that's because that is acutely uncomfortable acutely painful and coupled very closely with the slow flow I mean I just find that just so frustrating"

4.6 Discussion

The findings demonstrate a comprehensive patient-centered account of the symptoms, signs and impact of UAB, elicited from a large purposive sample of male and female patients with a primary diagnosis of DU. The patient reported experience of UAB has been revealed to be a complex myriad of storage and voiding symptoms. These findings provide the basis for essential evidence of content validity for the ICIQ-UAB.

The voiding symptoms of a slow stream, hesitancy and straining were reported by the majority (over 50%) of the patients in the sample. The flow rate was usually described as reduced, and in some individuals could be very severe and bothersome. A delay before urinating (hesitancy) could last for a few seconds to several minutes, dependent on the individual and the context. The symptom of straining (to initiate, maintain or finish urination) is of particular note as it was particularly well represented in the DU (only) group and is not currently included in the 2015 symptomatic definition [14]. These classic voiding symptoms are consistent with a weak bladder contraction and are in accordance with symptoms associated with DU in the literature [15,19,21,25]. Other voiding symptoms including intermittency, spraying (in men only), and urinations of a small volume per void were also elicited in several of the patients.

The storage symptoms reported by a large proportion of the patients included nocturia, increased daytime frequency, urgency and incontinence. These symptoms were key to the patient reported UAB experience, with a high prevalence of spontaneous reporting and often severe associated bother. Nocturia was the most commonly reported overall symptom as most patients described having to get up at least once in the night to urinate. Daytime urinary frequency was described by patients as being symptomatically bothersome in both low and high urinary frequencies. However, the higher prevalence and severity of incontinence in the groups with co-existing DO and SUI suggests that this symptom may be more related to these co-existing condition rather than underlying DU. The underlying aetiology of nocturia is also known to be complex, and may be more often a consequence of health or behavioural factors unrelated to lower urinary tract dysfunction [47,48].

A number of postmicturition symptoms including a sensation of incomplete emptying, the need to immediately re-void, and dribbling were frequently reported in the sample. A few patients described a perceived reduction in sensation of the fullness of the bladder, which also contributed to their feeling of suspicion of incomplete bladder emptying. A minority of

patients perceived an association with bowel issues and the severity of their symptoms. A feeling of incomplete emptying, absent or reduced sensation, bowel straining and a feeling of incomplete bowel emptying has been linked to DU patients by a recent database study of DU patients [21]. The majority of the participants had a post void residual (77% had a PVR of >30ml). Perhaps as a result, there was a high proportion of patients who were historically or currently self-catheterising and experienced UTIs in the past, including some who had experienced acute retention episodes. Lower urinary tract pain was reported by a minority of patients but the accounts were variable with regard to the type, source and level of pain experienced, suggesting a number of possible aetiologies.

Previous qualitative research into the impact of LUTS supports the findings that there can be a broad impact on patient's lives associated with these symptoms [9,49,50]. Disruption to sleep due to waking several times in the night and the lifestyle inconveniences caused by increased daytime urinary frequency were particularly apparent in the current study. The requirement to plan ahead around the location of toilets, detrimental effect on social life, embarrassment in certain situations and knock-on effect on self-esteem and confidence are supported by other qualitative studies in male patients with LUTS [50,51]. The inconvenience of self-catheterisation and impact of urinary tract infections also had a significant effect on quality of life for some patients. Incontinence, although more associated with patients with DU and co-existing DO or SUI, was often very bothersome with an associated effect on aspects of quality of life. However, many of the UAB patients also felt they were able to manage their symptoms to minimise the impact on their lives, as also documented in Glover et al (2004).

The necessity for the development of non-invasive tests for DU has been highlighted by previous studies [15,20]. The distinction between DU and BOO is of particular interest, as corrective surgery for voiding LUTS may be of little benefit for DU patients [203,204]. The current study did not identify any clear unique symptoms that may be attributed to UAB and the presence of co-existing urological conditions and associated overlap of symptoms in the sample further complicates diagnosis. The ICIQ-UAB instrument is intended to be an outcome measure rather than a diagnostic tool. However, recent research which showed differences in relative prevalence of symptoms in DU and BOO patients is encouraging for further research into the development of scoring algorithms to detect DU severity [21]. The combined use of the PRO tool and other non-invasive techniques such as ultrasonic measurement of the detrusor wall thickness [77] may also aid diagnosis.

Strengths and limitations

A strength of this study was that all patients were clinically verified to have a primary diagnosis of DU by urodynamics, confirming their likely status as patients with UAB. Although a qualitative study design has many advantages, it does not include the corroboration of reported findings with objective clinical measures (e.g. bladder diaries). For this reason, any indications of prevalence or severity of symptoms should be interpreted within the appropriate subjectivity of their context. Indeed, the study is not intended to produce representative epidemiological data but to elicit the overall patient experience of UAB. Although the patients interviewed were all White British, further interviews are described later in the thesis with patients from the US and Japan. These will explore potential differences in how patients describe the UAB experience from other cultures and ethnicities. Any further concepts elicited as a result of these interviews will increase our understanding of the condition and inform culturally adapted versions of a PRO instrument for its assessment.

4.7 Development of the draft ICIQ-UAB v1

The qualitative interviews represented progress in our understanding of how the clinical diagnosis of DU manifests as symptoms, by a thorough exploration of the lived experience of patients. This knowledge informed the following update of the preliminary conceptual framework (Figure 4) and the generation of the items included in the initial draft ICIQ-UAB (version 1).

4.7.1 Updated conceptual framework

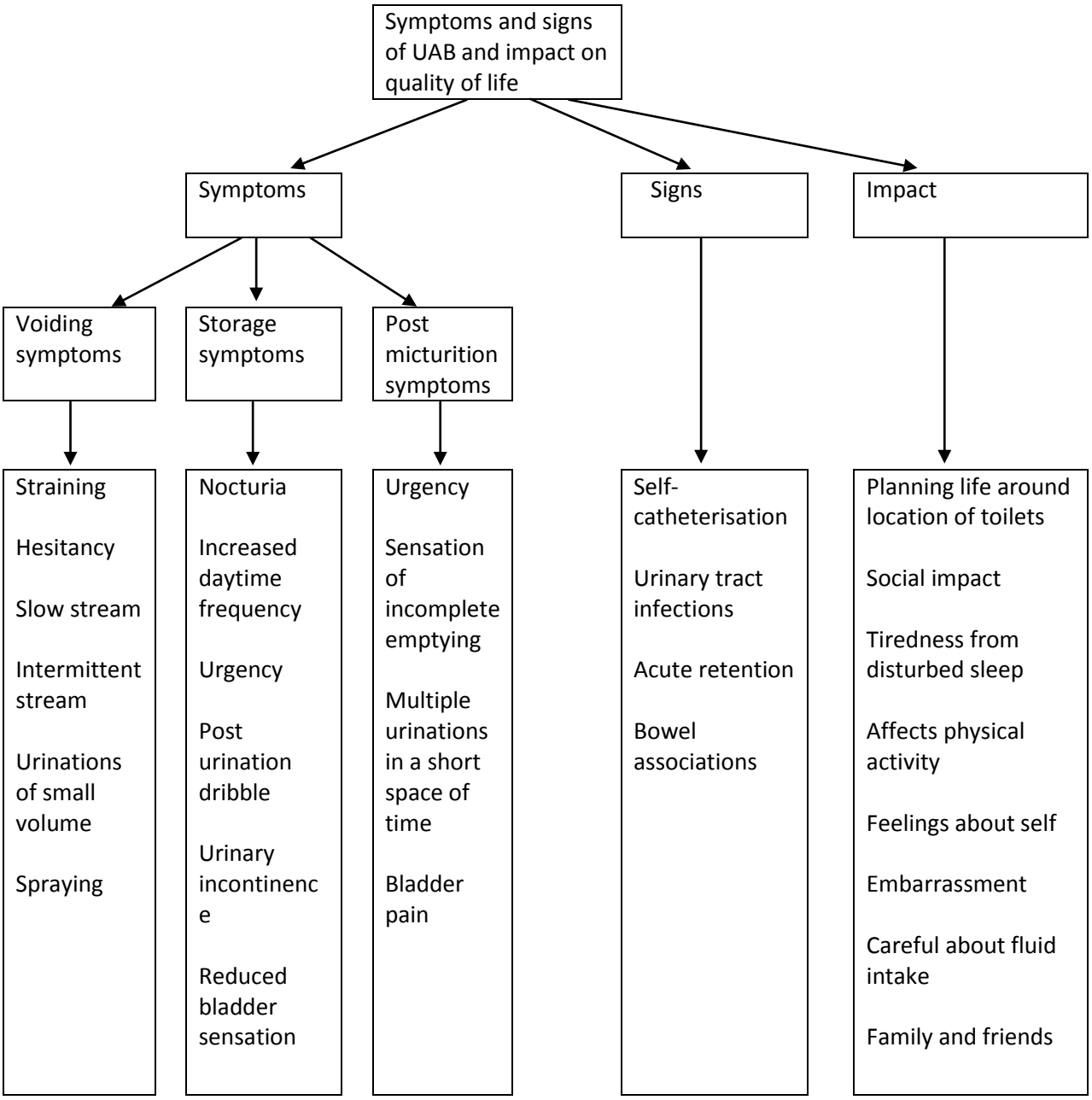


Figure 4. Updated conceptual framework.

4.8 Instrument development and item selection

Layout

The layout of the initial draft ICIQ-UAB was based on the template of existing ICIQ modules. These have been designed to be simple and easy to read, with items that are short, with clear response options [199]. The layout and facility of understanding will be assessed further at the cognitive interview stage and any necessary changes made.

Recall period

The recall period was selected as 24 hours for the symptom questions, which reflected the necessity for a specific time period and to facilitate the ability of patients to accurately recall their symptoms. The recall period was selected as 'over the last four weeks' for the impact questions. These recall periods are subject to change following further evidence at the cognitive interview stage.

Items

Decisions were systematically taken regarding the importance of inclusion of items, including their content, language and response items. A concept was deemed sufficiently important for item inclusion if it was considered the following:

- Well represented from the qualitative evidence in the DU only group.
- Deemed essential or desirable for inclusion by the CRP.
- Mentioned spontaneously by patients.

The following were considerations that reduced the importance of the inclusion of an item or multiple items, to cover a particular concept:

- More qualitative evidence for representation in the groups with DU and co-existing urological conditions (e.g. incontinence) than in the DU only group.
- Not mentioned spontaneously by patients.
- Deemed not essential for inclusion by the CRP.

The choice of wording for each item and their response items was devised using the following considerations:

- The idiomatic use of language from the patient accounts to ensure specificity to the UAB population.
- Existing validated ICIQ module questions that covered the same concept. For some items, these were used as a starting point before the cognitive interview stage allowed iterative revisions to improve their relevance and specificity to the UAB population.
- ICS definitions of the target symptom.

Table 13 summarises the evidence on which items were included.

Table 13. Summary evidence for inclusion of concepts for ICIQ-UAB v1.

Concept/symptom	DU (n=19)	DU & DO (n=8)	DU & SUI (n=7)	DU & BOO-E (n=5)	DU & BOO (n=5)	Was this reported spontaneously in DU group?	Majority CRP consensus	Which existing ICIQ modules include related items?	Number of questions included on this concept in ICIQ-UAB v1	Question number(s) in ICIQ-UAB v1
Nocturia and/or nocturnal voids	15	8	5	4	2	Yes. Strongly mentioned by several.	Essential/ Desirable	ICIQ-MLUTS ICIQ-Nqol	1	7
Increased daytime frequency	13	6	0	2	3	Yes. Strongly mentioned by several.	Essential	ICIQ-MLUTS	1	3
Straining	14	2	3	2	2	Yes	Essential	ICIQ-MLUTS LF ICIQ-FLUTS	4	12, 13, 14, 15
Hesitancy	13	2	3	3	3	Yes. Mentioned by several men.	Essential	ICIQ-MLUTS	2	4, 5
Slow stream	13	3	5	4	3	Yes by many	Essential	ICIQ-MLUTS ICIQ-MLUTS LF	2	18, 19
Urgency	11	3	3	3	2	Yes by a few	Essential/ Desirable	ICIQ-MLUTS	1	16
Intermittent stream	11	2	2	0	2	Not specifically	Essential	ICIQ-MLUTS	1	20
Sensation of incomplete emptying	8	1	4	4	2	No	Essential	ICIQ-MLUTS	2	8, 9
Need to immediately re-void	12	1	3	1	2	Yes by several	Not consulted	ICIQ-MLUTS LF	1	24
Reduced sensation of bladder fullness*	2	0	1	2	0	Yes	Not consulted	None	0	

Concept/symptom	DU (n=19)	DU & DO (n=8)	DU & SUI (n=7)	DU & BOO-E (n=5)	DU & BOO (n=5)	Was this reported spontaneously in DU group?	Majority CRP consensus	Which existing ICIQ modules include related items?	Number of questions included on this concept in ICIQ-UAB v1	Question number(s) in ICIQ-UAB v1
Urinations of small volume	4	0	1	1	1	Yes	Not consulted	None	1	26
Post micturition dribble	7	4	0	3	3	Very little mention	Essential/ Desirable	ICIQ-MLUTS	1	21
Urinary incontinence	5	4	6	1	4	Very little mention spontaneously	Desirable	ICIQ-MLUTS ICIQ-UI SF	1	6
Lower urinary tract pain	6	3	0	1	1	Not specifically	Not consulted	ICIQ-FLUTS	2	22, 23
Clustering of urinations*	6	3	1	1	0	No	Not essential	None	0	
Self-catheterisation	10	3	1	4	5	No	Not consulted	None	1	11
UTIs	8	2	3	3	1	Yes particularly by women	Not consulted	None	1	17
Retention	4	0	1	0	1	No	Not consulted	ICIQ-MLUTS LF	1	10
Temporarily unable to pass urine*	4	0	0	0	0	No	Not consulted	None	0	
Spraying	4	0	0	5	0	No	Not consulted	None	1	25
Associated bowel symptoms	3	1	1	2	1	No	Not consulted	ICIQ-B	1	36
Impact										
Planning life	13	7	3	2	2	Yes by many	Desirable	ICIQ-OABqol	1	27

Concept/symptom	DU (n=19)	DU & DO (n=8)	DU & SUI (n=7)	DU & BOO-E (n=5)	DU & BOO (n=5)	Was this reported spontaneously in DU group?	Majority CRP consensus	Which existing ICIQ modules include related items?	Number of questions included on this concept in ICIQ-UAB v1	Question number(s) in ICIQ-UAB v1
around location of toilets										
Social impact	4	2	1	1	0	Yes	Desirable	ICIQ-OABqol ICIQ-LUTSQoL	1	28
Tiredness from disturbed sleep	7	2	1	1	1	Yes by several	Essential/ Desirable	ICIQ-Nqol	3	29, 30, 31
Affects physical activity	3	2	1	1	0	Yes	Desirable	ICIQ-LUTSQoL	1	32
Feelings about self	3	1	5	2	3	Yes	Desirable	ICIQ-OABQoL	1	33
Embarrassment	4	4	2	0	1	Yes	Desirable	ICIQ-B ICIQ-LUTSQoL	1	34
Careful about fluid intake	8	3	2	1	1	Yes	Not consulted	ICIQ-LUTSQoL	1	35
Sex life	1	1	2	0	0	No	Not consulted	ICIQ-LUTSQoL	0	
Family and friends	5	1	2	2	0	No	Not consulted	ICIQ-LUTSQoL	0	

Abbreviations: Detrusor underactivity (DU), detrusor overactivity (DO), stress urinary incontinence (SUI), equivocal obstruction (EO), bladder outlet obstruction in the equivocal range (BOO-E), bladder outlet obstruction (BOO), clinical review panel (CRP), International consultation on incontinence questionnaire (ICIQ), ICIQ-underactive bladder (ICIQ-UAB), ICIQ-male lower urinary tract symptoms (ICIQ-MLUTS), ICIQ-nocturia quality of life (ICIQ-Nqol), ICIQ-MLUTS long form (ICIQ-MLUTS LF), ICIQ-Urinary Incontinence short form (ICIQ-UI SF), ICIQ-female lower urinary tract symptoms (ICIQ-FLUTS), ICIQ-bowels (ICIQ-B), ICIQ-overactive bladder quality of life (ICIQ-OABqol), ICIQ-lower urinary tract symptoms quality of life (ICIQ-LUTSQoL).

*Question included in later versions of ICIQ-UAB during cognitive debriefing stage.

4.9 Conclusion

The concept elicitation sub-study represents progress in our understanding of how the clinical diagnosis of DU manifest as symptoms, by account of the lived experience of patients. The evidence from the literature review, expert clinical panel and the concept elicitation interviews were used as the basis for the first draft of the ICIQ-UAB (Appendix 5). This draft version represents a comprehensive item list of the concepts which are deemed relevant by patients.

Following further testing and refinement, the resulting PRO instrument for the assessment of UAB is potentially important for the assessment of UAB patients, when used alongside other non-invasive methods. This knowledge of the patient experience of UAB is also valuable to the further development of the 2015 symptomatic based definition.

Chapter 5 Cognitive interviews

5.1 Introduction

The cognitive theory model on which cognitive interviewing is based consists of four stages to explain how questionnaire items are processed and answered by subjects. Respondents must first understand the question, then recall the item-specific information, assess the type of information required and finally decide on the response [147]. Cognitive interviews follow a 'think-aloud' methodology, which ask respondents to verbalise the cognitive process they used to interpret and answer each of the survey items [115,141]. The cognitive interview process is essential to demonstrate evidence of content validity, patient comprehension of the draft measure, and to make appropriate modifications to the instrument.

5.2 Aims and objectives

The purpose of the cognitive interviews was to meet the following objectives:

- To ensure the items are understood and interpreted as intended by patients, using appropriate and patient-centered language.
- To select appropriate and clear response items to each question with similar and well-spaced intervals, which reflect the full range and attributes of the concept measured.
- To select a recall period appropriate to the instrument's purpose and intended use, whilst considering the variability and frequency of the concepts measured and characteristics of the condition.
- To ensure the format, layout, order and sequence of items maximise understanding, readability, and facility of completion [115].

5.3 Methods

5.3.1 Site selection and recruitment

Patients were recruited at the BUI at Southmead Hospital in Bristol, Royal Hallamshire Hospital in Sheffield and at the Freeman Hospital, Newcastle in May 2015. The medical records of prospective participants were screened for their eligibility for this phase of the study. Those who met the inclusion criteria were sent a postal invitation, containing a letter introducing the

study, an information sheet and a form which allowed them to indicate their willingness to take part by return post. A follow-up call was made if there was no response from the invitee within two weeks. If the invitee was unable to take the call, then a voicemail message was left and one further call was made before the participant was deemed uncontactable. If a positive reply by post was received, the study researcher contacted the invitee by phone and negotiated a suitable time to conduct the interview.

5.3.2 Sample size and target population

Cognitive interview participants were recruited to be broadly representative of the target population in which the PRO instrument will be used (e.g. age, sex, ethnicity, socioeconomic status, literacy and characteristics of condition). Based on the complexity of the concepts underlying the draft UAB PRO, it was expected that 25–30 interviews would be required with 3–5 participants per round [124,141].

5.3.3 Study inclusion and exclusion criteria

The patient inclusion and exclusion criteria were the same as used for the concept elicitation study in chapter 4. All participants had a diagnosis of detrusor underactivity following urodynamics.

5.3.4 Ethics and informed consent

This study was conducted with favourable opinion from the Southmead Research Ethics Committee (now the Bristol South Research Ethics Committee) for ICIQ questionnaire development activities: reference 087/99. Before each interview, the participants were reminded of the purpose of the interview, the procedure on confidentiality, and were given the opportunity to ask questions. Participants signed an informed consent form in-person before each interview.

5.3.5 Conduct of interviews

A cognitive interview guide was developed to allow for the retrospective (after the patient has completed all items on the questionnaire) debriefing of the instrument (Appendix 6). This included general readability and layout, length, completion instructions, recall period, item content, item clarity and wording, bother items and general comments about item inclusion. A practice interview with one participant allowed familiarisation with the guide and to address potential issues that could occur in the interview. Successive rounds of cognitive interviews

were then scheduled to determine whether the target population understood and interpreted the content of the first version of the UAB PRO as intended. Interviews were conducted in a room which minimised the risk of interruption and maintained the privacy of the participant. Following consent, participants were asked to complete the questionnaire in its entirety with minimal input from the interviewer before any probing questions were asked. The audio recorder was switched on and subjects were asked to make general comments on the questionnaire length, layout, coverage and facility of completion. They were then asked about the clarity of the completion instructions and their understanding of the recall period. For each item, systematic probing questions were asked which allowed the assessment of the subjects' comprehension in relation to the intended meaning. Subjects were encouraged to 'think aloud' and to be as 'open and honest as you like when you tell me your thoughts and comments'. Notes were made by the interviewer which summarised the subjects' responses and were documented throughout using the boxes corresponding to each item in the interview guide.

5.3.6 Data management

Following each interview, the audio recording was transcribed verbatim. All files were password protected and stored on the NBT secure server. All participant data was anonymised using a unique study identifier.

5.3.7 Item tracking matrix

After each round of interviews, changes were made to the draft PRO instrument according to patient feedback consensus. Any changes made, and the rationale for these, were documented and described in item tracking matrices prepared in an Excel spreadsheet. A summary of each subject's comments, as well as any input or recommendations from the clinical panel or other sources was recorded here. This collated the relevant information to facilitate informed decisions during the item development. Changes were only made if several patients within a round had the same misunderstanding or made a similar recommendation for improvement. This avoided changes being made on the basis of a single patient. The responsibility for decisions made on the revision of items after each round was ultimately by the author. Additional rounds of cognitive interviews were scheduled until no further changes were required.

5.4 Results

5.4.1 Cognitive interview participant characteristics

A total of 36 participants (nine female, 28 male, age range 26–88 years) were interviewed in-person over 10 successive rounds of semi-structured interviews in March and April 2015. A final eleventh round of interviews was conducted by phone to test some final minor changes with five previous participants. Twenty eight patients were interviewed at Southmead hospital, four at Royal Hallamshire Hospital, Sheffield, and five at the Freeman Hospital, Newcastle. Twelve of the participants had previously been interviewed during the concept elicitation interviews. All were White British, and came from a variety of different educational backgrounds and occupations. Table 14 shows the sample population diagnostic groups and demographic characteristics.

Table 14. Summary of demographic and clinical characteristics.

Clinical or demographic characteristic	**Total sample (n=36)**
DU DU + DO DU + SUI DU + BOO-E DU + BOO DU + Mixed	15 5 1 9 2 4
Males n (%)	27 (75)
Females n (%)	9 (25)
Age [years] median (IQR)	68 (14)
ISC current or historical n (%)	11 (31)
PVR >0ml n (%)	18 (50)
PVR [ml] median (IQR)	90 (184)
BCI (males only) median (IQR)	73 (17)
BOOI (males only) median (IQR)	20 (19)
$p_{det}Q_{max}$ [cmH_20] median (IQR)	33 (20)
Q_{max} [ml/s] median (IQR)	7.5 (6)
BVE (%) median (IQR)	79 (49)

Abbreviations: Detrusor pressure at maximum flow ($p_{det}Q_{max}$), maximum flow rate (Q_{max}), bladder contractility index (BCI) calculated by BCI=$p_{det}Q_{max}$+5Q_{max}, Bladder Outlet Obstruction Index (BOOI) calculated by BOOI=$p_{det}Q_{max}$-2Q_{max}, Post Void Residual (PVR).

5.4.2 Instructions

The purpose of the instructions was to provide a brief introduction to the questionnaire. The original instructions stated:

> "Many people experience urinary symptoms some of the time. This questionnaire aims to find out whether you experience symptoms known to be associated with underactive bladder, and whether or not these symptoms have an impact on your everyday life. We would be grateful if you could answer the following questions, thinking about how your symptoms have been, on average, over the **LAST 24 HOURS.**"

The wording and clarity of the instructions were reported to be clear by all subjects and did not require any significant changes in content. The final instructions were as follows:

> "Many people experience urinary symptoms some of the time. This questionnaire aims to find out whether you experience symptoms associated with underactive bladder, and whether or not these symptoms have an impact on your everyday life. Please answer the following questions thinking about how your symptoms have been over the LAST 24 HOURS."

5.4.3 Layout

The initial layout and formatting were based on the template of the existing ICIQ modules . All subjects found the layout to be clear and easy to read, so no changes were made to this format. An example item is shown in Figure 5.

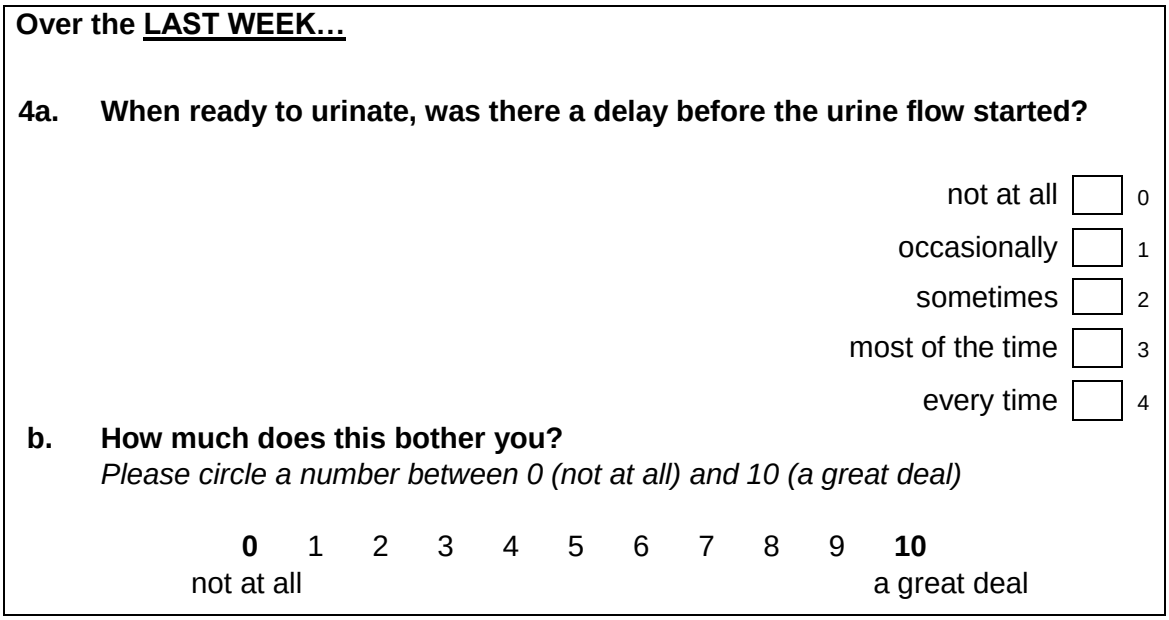


Figure 5. Example item.

5.4.4 Bother

Each item has the associated question 'how much does this bother you?', which allows the subject to indicate their assessment of this on a 0–10 numeric rating scale. All subjects in the cognitive interviews were able to answer the question as intended. Two subjects suggested that they preferred the question to have worded response options (e.g. not at all, slightly, somewhat, very much etc.) but did not regard this as essential. A few subjects answered the associated bother question with respect to how they would feel if they had the symptom in the future. However, this did not affect how they answered the bother questions for other items where the symptom was present.

5.4.5 Recall period

The selection of the appropriate recall period depends on the period over which a patient can accurately recall the event. The recall period of version 1 was based on a 24-hour recall period for the symptom items and a 4-week recall period for the impact items. This underwent several revisions during the cognitive interview rounds in to test whether a 4-week, 1-week, or

24-hour recall period was the most appropriate for the symptoms and impacts to be measured in the study population.

Some patients favoured the 24-hour recall or the 1-week recall period, while others reported that the recall period was too short or too long. For example, a patient who preferred a shorter recall period stated: “I was thinking of a day a normal day”. Another patient (CD P2) stated: “The questions seemed to come at come at me in a general way you know unless it specifically says and you know in like the last 4 weeks or so”. Many of the patients reported that they had not seen the recall period or ignored it. After round two, the recall period was put as a reminder at the top of each page to encourage subjects to adhere to the recall period provided. However, due to continuing issues with subjects not reading or ignoring the recall period, a reminder of the recall period was inserted for each question. Two versions of the questionnaire were tested and retained from the eighth round of cognitive interviews. One version has a 24-hour recall period and only contains the symptom questions. The other has a recall period of over the ‘last week’ for symptom as well as impact questions.

5.4.6 Response options

The response options were tailored to the item in question and are discussed where relevant to the item. However, many required more general response items. The initial response options for these items were ‘never, occasionally, sometimes, most of the time, all of the time’. After round 1, the response options were changed to ‘never, occasionally, sometimes, often, every time’. It was noted by several subjects that the response options did not encourage them to think about a short recall period of 24 hours. After round 4, alternative response options were designed for a shorter recall period: ‘not at all, once, a few times, many times and every time’. However, following feedback from patients in rounds 6 and 7, it was clear that different response options were required for different recall periods. The final response options were introduced in the eighth round of interviews and in each version thereafter with a ’24-hour/Last week’ recall period. The respective options were ‘not at all/not at all, on occasion/occasionally, about half the time/sometimes, most times/most of the time, every time/every time’.

5.4.7 Symptoms

Daytime urinary frequency

The original item was ‘How often do you pass urine during the day, on average? (1–6 times, 7–8 times, 9–10 times, 11–12 times, 13 or more times)’. An important alteration to the response options was to include a frequency option of 1–3 times, as this reflected the infrequent number of urinations by some UAB patients. The final item for the version with a 24-hour recall period was ‘How many times did you urinate during the day? (1–3 times, 4–6 times, 7–9 times, 10–12 times, 13 or more times)’ and for the version with a 1 week recall period: ‘During the day, how many times did you urinate, on average? (1–3 times, 4–6 times, 7–9 times, 10–12 times, 13 or more times)’.

Hesitancy

The original item was worded ‘Is there a delay before you can start to urinate? The words ‘when ready to urinate’ were added to the item stem to improve the clarity of circumstance in which the question was being asked. The final item was ‘When ready to urinate, was there a delay before the urine flow started?’

Need to concentrate to start void

The original item was ‘How often do you feel like you have to concentrate to start urinating?’. Patients understood the item as intended, for example CD P2 reported: “*your brain's telling you you've got to go and then for some reason you can't go and it's a battle then within the mind”*. At the seventh round the item was changed to its final iteration: ‘When ready to urinate, did you feel you had to concentrate to start urinating?’

Urinary incontinence

The original item stem was ‘Do you ever leak urine before you can get to the toilet?’. Patients mentioned different types of incontinence, for example after exercise or when coughing or sneezing CD P21: *“I think once I've got a cold and a cough then I often [leak]”*. The item modified to give examples of circumstances of leakage ‘before you could get to the toilet, or when you cough or sneeze’. This clarified the intention of the question and to distinguish this item from the question on post-micturition dribble. The final item was ‘How often did you leak

urine (e.g. before you could get to the toilet, when physically active, or when you coughed or sneezed)?'.

Nocturia and/or nocturnal voids

The original item was interpreted as intended: 'During the night, how many times do you have to get up to urinate, on average? (none, one, two, three, four or more)'.

Reduced sensation of bladder fullness

The original item was 'Would you say that your sensation of how full your bladder is… (not reduced, a little reduced, reduced, very reduced, extremely reduced)'. Following patient feedback this item was reworded for clarity of language and replaced with a question which aimed to establish how well respondents could detect fullness of their bladder. For example, CD 26: *"Yeah it's very difficult to know how full your bladder is, although I have some sort of feeling there"*. The final item was 'Did you find it difficult to tell how full your bladder was? (not difficult, a little difficult, difficult, very difficult, and extremely difficult)'.

Sensation of incomplete emptying

The original item was 'How often do you feel like your bladder has not emptied properly after you have urinated? Patients understood the item as intended, for example, patient CD P2: *"you know, you're out with company and you fancied a toilet and you think, have I emptied?"*. However, subsequent cognitive interviews showed that subjects who were either aware or unaware of a residual would respond positively to this question i.e. those who could feel it or those who were just 'suspicious' there was still a residual due to other cues. The item was changed to 'After you have urinated, how often do you have a sensation that your bladder is not completely empty?'. The word 'soon' was added to the item to further improve clarity around context. The final item was 'Soon after you had urinated, how often did you have a sensation that your bladder was not completely empty?'.

A second item was included from patient feedback that the item on the length of time spent in the bathroom was not covered by the slow stream item. The included item was interpreted as intended: 'How often did you wait a bit longer in the bathroom after urinating, to make sure your bladder was as empty as possible?'.

Acute retention

The original question was worded: 'Have you ever blocked up completely so that you could not urinate at all and had to have a catheter passed to drain the bladder?' Two patients in round 2 interpreted the question to mean self-catheterisation. This was addressed by changing the order of the question to after the self-catheterisation item to improve the distinction between these items. Adding 'and had to go to hospital' also improved the clarity of context. The final question was worded, 'Have you ever been unable to urinate at all and had to go to hospital to have a catheter tube inserted to drain the bladder?'.

Self-catheterisation

The item was worded 'How often do you need to self-catheterise?' with response options 'never, less than once a month, a few times a week, once a day, every time'. The main modification to the original question was to establish the most appropriate response options based on patient feedback. Incorporating an item-specific recall period of 'over the last month' was considered necessary as some subjects may have tried self-catheterisation in the past. The final wording of this item was 'Over the last month, how often did you self-catheterise (not at all, about once a week or less often, two or three times a week, about once a day, more than once a day, every time you urinated)'.

Straining

The original question was worded: 'Do you have to strain or squeeze to urinate? (never, occasionally, sometimes, most of the time, all of the time)'. The words 'squeeze' and 'strain' seemed to be contradictory, for example, CD P2: *"I don't know about squeeze I mean straining it's the force to balance to try and you know to activate…"*. The question also appeared to be unclear as to when during urination straining occurred. Three items were tested that probed more specifically about whether straining was involved in initiating, maintaining and the end of urination. The first asked 'Do you have to strain to start urinating? (never, occasionally, sometimes, most of the time, all of the time)'. This was more relevant CD P16: "*Yeah I mean to force it to force your muscles and trying to make something happen"*. The final item was 'How often did you strain to start your urinations? The second item 'Do you ever strain to try and improve the flow of your urination', although relevant to a minority of patients, was removed after the third round of cognitive interviews due to the perceived overlap with the other items related to straining. The third straining item tested captured the frequency with which the

patient strained to finish urination: ‘How often did you strain towards the end of your urinations to try and empty your bladder?’

Urgency

This item was intended to measure the frequency that a patient experiences urgency (a sudden compelling desire to pass urine which is difficult to defer). The original question was ‘How often do you experience a sudden desire or urge to urinate which you are unable to ignore?’. Patients interpreted the question as intended with one respondent (CD P2) saying: *“absolutely must go feeling”*. The words ‘urge’ and ‘desire’ were explored in relation to this concept but following patient feedback from the cognitive interviews and input from the CRP ‘a sudden or strong need’ was confirmed as eliciting the intended response in the subjects. The addition of ‘and had to rush to the bathroom’ further clarified the context of the question. The final item was ‘How often did you experience a sudden or strong need to urinate which you were unable to ignore, and had to rush to the bathroom?’.

Urinary tract infection

The original item was ‘Have you had a urinary tract infection in the past month? (Yes, No)’. Patients could be unsure when they had their last UTI, or if in fact it was a UTI, such as patient CD P28: *“You know and in the last year, I don’t think it is I think it’s over a year the last time”*. An ‘unsure’ option and clarification ‘for which you took medication’ improved understanding. The final item was ‘Over the last 12 months, have you had a urinary infection for which you took medication? (no, unsure, once, twice, three or more times)’.

Slow stream

The original item was worded ‘Would you say the strength of your urinary stream is: normal, occasionally reduced, sometimes reduced, reduced most of the time, reduced all of the time’. This was generally interpreted as intended but subjects were more concerned with the severity of their slow stream than the frequency of the symptom. For example, CD P6 stated: *“it was extremely reduced, it would be like dripping out.”* This question replaced with ‘Would you say the flow of your urinary stream is: not reduced, a little reduced, reduced, very reduced, extremely reduced. The words 'Strength of' were also added before ‘flow’ in this round. ‘Normal’ was added to the first response option to give the respondents some reference for comparison. The final item was ‘On average, would you say that the strength of

flow of your urinary stream was… (Normal (not reduced), a little reduced, reduced, very reduced, extremely reduced)'.

The second item was introduced when recommended as a potential question by a member of the clinical panel: 'How long do you stay in the bathroom to finish urinating? (less than a minute, 1–5 minutes, 6–10 minutes, 11–15 minutes, more than 15 minutes)'. After some rewording to improve clarity, the final item was 'What was the longest time that you needed to spend in the bathroom to finish urinating? (less than a minute, 1–5 minutes, 6–10 minutes, 11–15 minutes, more than 15 minutes)'.

Intermittency

The original item was 'Do you stop and start more than once while you urinate?'. Patients understood the item as intended, for example CD P2: *"it stops… and then you've got to wait for it to come on stream again"*. The reversal of the order of the words 'stop' and 'start' was considered and tested but ultimately the original order was maintained as it was more colloquial. The final item was 'How often did you stop and start more than once, during your urinations?'.

Post-micturition dribble

The original item was worded 'How often does a few more drops leak out into your underwear shortly after you have finished urinating and have dressed yourself? This question was understood as intended. For example, patient CD P6 stated: *"well it can be a bit embarrassing as though you've wet yourself but like I say it's never it was never bad"*. The final item was 'How often did a few drops leak out into your underwear shortly after you had finished urinating and had dressed yourself?'.

Bladder pain

The original item was 'Do you experience pain in your bladder?'. The item was removed in the third round of cognitive interviewing as there was no consensus on the type, location or circumstance. Pain also did not appear to be a ubiquitous problem either in the concept elicitation or the cognitive interviews. Several also interpreted the question as their normal 'discomfort' when their bladder was feeling full.

Need to immediately re-void

The original item was 'Do you have to urinate again (within 15 minutes) after you thought you had finished urinating?'. After patient feedback, for example CD P3: *"I can't say it's within 15 minutes or anything like that it could be half an hour"* and patient CD P17: "*for me I think maybe the less it could be is half an hour later"* the question was reworded for clarity around the length of time. The final question was 'After you had urinated, how often did you have to return to the bathroom to urinate again, within a short space of time (e.g. within 15 minutes)?'.

Temporarily unable to pass urine

This item was introduced in round 2 of the cognitive interviews and was originally worded 'How often have you been unable to urinate when you would like, and had to come back later?'. The aim was to capture those who cannot voluntarily urinate unless feeling urgency as well as those who sometimes find themselves unable to voluntarily urinate independent of desire. The wording was reworded for clarity around circumstance: 'How often did you go to the toilet to pass urine but were unable to urinate <u>at all</u>, so had to return to the bathroom to try again later?

Spraying

The original item was worded 'How often do you feel like you have no control over the direction of your stream?'. Although several of the men reported lack of control over the direction of the stream, this was primarily due to poor flow, for example CD P8: *"I can't control, I can't control the strength"*. It was also considered less relevant to women due to their sitting down position whilst urinating. The question was deemed the lowest priority by the clinical panel so was removed after the second round of interviews.

Small volume of urine per void

The item was originally worded, 'How often do you feel like you were not able to pass what you might consider a satisfactory amount of urine?'. There was some feedback from patients that the question was too subjective due to there being no point of comparison. However, it was acknowledged that for subjects for whom this was a symptom, the item was interpreted as intended, for example CD P10: *"yeah that's what I am saying to me the volume is not great anytime"*. The final item was 'How often were you only able to pass a small volume of urine?'

Associated bowel symptoms

The original question was worded: 'Are your urinary symptoms affected by your bowel movements?'. The intent of the item was to measure the frequency with which a patient experiences an association between bowel problems and the severity of their urinary symptoms. The question was reworded in round 4 to improve clarity because of the many different ways that bowels can interact with the urination process. The final item was 'Did you have problems with your bowels? (no, yes) If so, were your urinary symptoms made worse by this?'

Clustering of symptoms

This item was included in round 7 due to feedback from the cognitive interviews and concept elicitation evidence that some subjects experience more severe symptoms at a specific time of the day or night. The original question was worded 'Were your symptoms worse at particular times? Please tick those that applied… (no, in the morning, in the afternoon, in the evening, at night)'. It is acknowledged that the specific symptom or symptoms which the respondent may be reporting is not known; nevertheless the wording of the question was confirmed as being interpreted as intended. The final item was 'Were your urinary symptoms worse at particular times? Please tick those that apply… (no, in the morning, in the afternoon, in the evening, at night)'.

5.4.8 Impact items

Planning around toilets

The original item was worded, 'How much do you have to plan your life around the location of toilets? (never, occasionally, sometimes, most of the time, all of the time)'. The patients interpreted the item as intended and gave feedback that this item was both relevant and important. For example, CD P3: *"if I do go out if I go out to the theatre or cinema it's a basic question of scanning the area to see where the toilet is".* The only change was the replacement of the words 'How much' with 'how often' for consistency with the other questions and the change to past tense. The final wording of the item was 'How often did you have to plan your life around the location of toilets?'

Social life

The original item was worded 'How often do you feel that your urinary problem interferes with your social life?'. Patients interpreted 'social life' as intended, for example, CD P17: *"relaxing time enjoying time with friends".* The final wording of the item was 'How often did you feel that your urinary symptoms interfered with your social life?'

Sleep

The original item was 'How often do your symptoms prevent you from getting the amount of sleep you needed?'. After round 1 'urinary' was added to precede 'symptoms' for clarity. After the second round the item was considered to be more concisely captured by the more general nocturia impact question, described below. As the item was also deemed the lowest priority by the clinical panel, it was not included in the questionnaire.

Tiredness

The original item was worded 'How often do you feel tired the next day because of having to get up at night to urinate?'. As above, the general nocturia impact question below was retained in favour of this item due to its increased relevance.

Impact of nocturia and/or nocturnal voids

The item was originally worded as 'Overall, how much impact does getting up at night to urinate have on your day to day life?'. After round 1, 'at night' was underlined for emphasis. In round 4, following feedback from the patients, 'impacts' was changed to 'affects' to improve comprehension. The item was interpreted as intended, for example CD P12: "*it does impact because at the end of the day I'm awake when I would be asleep"*. The final item was 'How often did you feel getting up <u>at night</u> to urinate affected your day to day life?'

Physical activities

The item was originally worded as 'How often do you feel that your urinary problem limits your physical activities (e.g. exercise, sport)?' Many reported that doing just 'sport' was not applicable to them. For example CD P12: *"I would say any kind of walking, exercise, any kind of sports"*. Patients gave feedback that their physical activity was more 'affected' (not limiting their 'ability'), due to their reliance on being close to a toilet. The final item was 'How often did

you feel that your urinary symptoms affected your physical activities (e.g. walking, swimming, sport)?'

Feelings about self

The original question asked 'Does your urinary problem affect the way you feel about yourself?' Patients understood the item as intended, for example CD P16: *"Well does it make me feel any less confident than I would normally be?"*. Others mentioned anxiety or depression when considering how to answer this question; for example CD P12: "*yes, I personally suffer from depression, someone else might not"*. The final item was 'Did your urinary problem affect the way you feel about yourself?'

Embarrassment

The original item was worded 'Does your urinary problem cause you to feel embarrassed?'. Patients understood the item as intended, for example CD P12: *"Okay people don't realise that I have got problems and they just think oh he needs to go to the toilet, but it still don't change the fact that I might feel a bit embarrassed about it"*. The final item was 'Did your urinary problem cause you to feel embarrassed?'

Fluid intake

The original item of the question was 'Are you careful how much fluid you drink?'. After round 1, it was noted that the both the amount and type of drink are monitored by patients so the item was altered accordingly. The final item was 'Did your urinary symptoms cause you to be careful about how much or the type of fluid you drink?'

Overall impact

The original item was 'Overall, how much would you say your urinary symptoms interfere with your everyday life?'. The response frame remained an eleven point Likert scale ranging from 0-10. No changes to the item stem were required.

5.5 Discussion

The current chapter documents the refinements to the draft ICIQ UAB version 1 during the cognitive interviews. The subsequent ten versions underwent modification from the original version including the removal and addition of several items based on patient feedback and

input from the clinical panel. Areas for improvement were highlighted throughout the cognitive interview process, including suggested changes to wording, clarity, and content which were implemented and further tested. The resulting versions of the ICIQ-UAB reflect the input of both patients and clinicians. The following discusses specific items and aspects of the instrument which were highlighted by this study which can be explored by the further investigation of the instrument's psychometric properties.

The item 'How often did you feel you have to concentrate in order to start urinating?' was found to be highly relevant to many of the cognitive interview subjects as many subjects related to this concept of 'having to concentrate' in order to start passing urine. The inclusion of the specific terminology used by patients is considered to particularly capture the difference between the concept of concentrating and the physical symptom of hesitancy captured by an alternative item. Any overlap in underlying concept is possible to explore in the next stage of the development process.

The items which related to bladder sensation underwent some exploration. These included: 'Did you find it difficult to tell how full your bladder was?', 'Soon after you have urinated, how often did you have a sensation that your bladder is not completely empty?', 'How often were you only able to pass a small volume of urine?' and 'After you have urinated, how often did you have to return to the bathroom to urinate again, within a short space of time (e.g., within 15 minutes)?'. Respondents may consider whether they are able to predict the volume of urine left in their bladder, including whether they trust the accuracy of their bladder sensations post-urination when answering each of these items. These items share underlying concepts due to a post-void residual being ubiquitous in the UAB population but elucidate different contexts. Any overlap will be explored in the next stage of the development process.

The concept elicitation interviews and clinical panel provided evidence that for some subjects there is an association between bowel problems and urinary symptoms resulting in the inclusion of the item: 'Did you have problems with your bowels? Then 'if so, are your urinary symptoms made worse by this?'. However, the inclusion of this item was subject to some scrutiny during the cognitive interview process due to it lacking sufficient specificity to be well understood by subjects. Further expansion into a number of different items was not considered desirable as the primary focus of the PRO was urinary symptoms. The item will be considered again pending further study.

The item 'Were your symptoms worse at particular times?' was included as several subjects experience more severe symptoms at a particular time of the day or night. It is acknowledged that the specific symptom(s) which the respondent is thinking about are not specified. Further study will help decide the merit of including this question.

The impact questions overall required very little modification. The final version included one nocturia-related impact question rather than the initial three items that were debriefed. Furthermore, for many respondents there is a necessity to be in close proximity to a toilet, covered by the item 'How often did you have to plan your life around the location of toilets?' This fundamental inconvenience has a secondary impact covered by the following two items: 'How often did you feel that your urinary symptoms affect your physical activity (e.g. walking, swimming, sport)?' and 'How often did you feel that your urinary symptoms interfere with your social life?'. Although there is an overlap of an underlying concept, these additional items clarify the context which justifies their inclusion at this stage. These are also generally supported for inclusion as fundamental features of the quality of life definition in any context.

The cognitive interviews uncovered that in some cases, patients would answer the bother questions according to 'hypothetical bother' if they did not have a particular symptom, rather than 'actual bother' related to their experience of a symptom. Any analysis of bother in further pilot testing will exclude bother scores for items in which a symptom was not reported by a patient. An electronic version of the questionnaire will also address this limitation as a bother question would only be asked if patients experience the symptom in question.

An appropriate recall period for the population is important so the instrument has an interpretable score within the specific context of measurement. The current recommendations are that a shorter recall period is advised in order to improve recall accuracy and provide specific rather than general data [99,115]. However, although some patients preferred a longer recall period for symptoms, there is a delicate balance between responding to the data and current thinking in the field. Two versions of the ICIQ-UAB will therefore go on to further psychometric testing. One version has a recall period of the 'over the last week' which applies to all items, including the impact questions. The other version has a recall period of 'the last 24-hours' and only contains the symptom questions.

It is acknowledged as a possible limitation the subjectivity that a single researcher may bring when doing qualitative research. There may also be a difference between what an interview respondent says and what he or she actually means. Therefore, the reliability of the findings

was dependent on the continual reflexive appraisal and referral to the original transcripts, particularly during the analytical process. It is also acknowledged that the subjects of the study may be influenced by an overt presence of the researcher (e.g. the effect of a male researcher when eliciting sensitive topics) and why the findings should always be interpreted in context [138]. In addition, it is acknowledged that as prior knowledge of underlying urological theory derived from the literature is likely to have had an influence on the decisions made. For example there may be a danger that only evidence that agrees with preconceived ideas are identified [205]. For this reason, care was taken during the interviews and analysis to ask neutral and non-leading questions in order to elicit reduce the chance of bias in the participant's responses. It was also noted if something was not mentioned, as this may be equally as important as what is verbalised. It is acknowledged that using patients to revise the questionnaire who were involved in the elicitation of concepts could be construed as biasing the development of the items. However, the impact of including 12 interviewees from the concept elicitation interviews in the cognitive interviews was considered unlikely to have influenced the decisions made as the further 24 interviews, including in-depth interviews from different sites in Newcastle and Sheffield were carried out until all questions were completed easily and interpreted as intended.

5.6 Conclusion

The elicitation of concepts, development of a draft version of the instrument, the process of conducting cognitive interviews in the target UAB population and subsequent revisions of the instrument conclude the initial qualitative phase of the questionnaire development. The documentation of this process in the current report provides support for its content validity and aids determination its context of use. As a result of the cognitive interviews, the revised versions of the draft instrument are considered comprehensive and ready for the next phase of development, psychometric testing. Figure 6 shows the item concepts and 31-item pool.

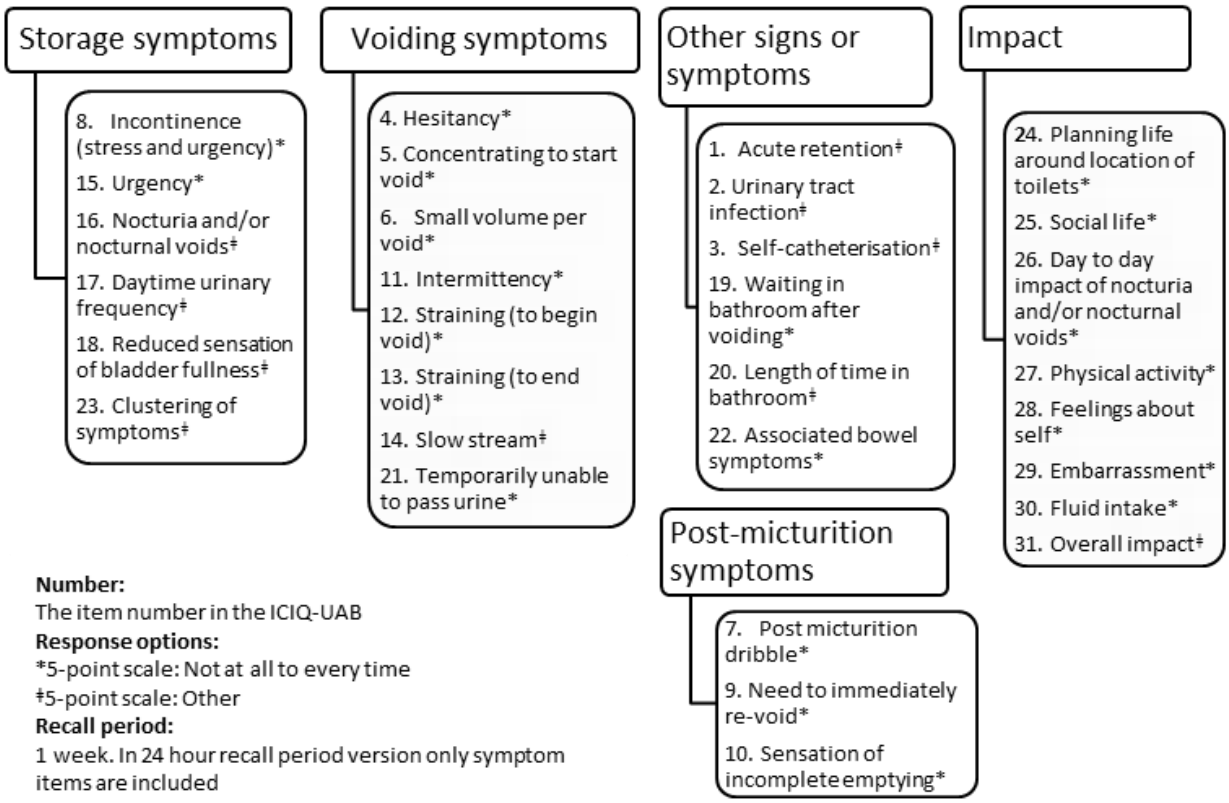


Figure 6. The 31-item pool for the draft ICIQ-UAB.

Chapter 6 Quantitative psychometric evaluation

6.1 Introduction

This chapter details the studies undertaken to assess the initial psychometric properties of the draft ICIQ-UAB, developed as a result of the qualitative phase detailed in chapters 4 and 5. This pilot study was designed for the initial quantitative exploration of the validity and reliability of the draft questionnaire in a new sample of patients diagnosed with DU. The aim was to provide additional evidence for decisions to be made surrounding the items to be included in the final questionnaire.

6.2 Methods

6.2.1 Study design

The study was designed to assess the preliminary psychometric properties of the draft ICIQ-UAB in the target population, using a standard test-retest design. The qualitative interview phase described previously, resulted in a developmental version of the ICIQ-UAB consisting of 31 items:

- Medical history and bother (3 items)
- Symptoms and bother (20 items)
- Impact and bother (8 items)

For the symptom items, a version with a 24 hour recall period was also tested, to allow comparison of the suitability of the recall periods. The two versions were as follows:

- ICIQ-UAB 1-week – all items with a 1 week recall period (31 items)
- ICIQ-UAB 24hrs – only symptom items, with a 24 hour recall period (20 items)

Participants were asked to complete a demographic and health information form (DHIF), both of the versions of the ICIQ-UAB, and concurrent instruments of known evidenced validity (detailed below) over a period of 10 days. A period between test-retest of approximately two weeks is considered reasonable to ensure the condition severity is not expected to change and there is sufficient time to reduce possible learning effects [103]. Table 15 details the content of each of the packages and the order in which each of the questionnaires were completed. All

questionnaires were compiled into a questionnaire booklet to direct the order of completion. Approximately half of the participants completed the ICIQ- UAB 1-week first at baseline (day 1) followed by the concurrent PROs and the ICIQ-UAB 24hrs at the end. For the other half of the participants the order of the UAB PROs was reversed to eliminate possible order effects of administration. The allocation to which PROM was administered was by subject ID, (odd numbers completed the ICIQ-UAB 1-week first).

Table 15. Study flow chart.

	Day 1	Day 2	Day 3		Day 8	Day 9	Day 10
DHIF	x						
ICIQ -UAB 1-week	x				x		
ICIQ-MLUTS/FLUTS	x						
ICIQ LUTSqol	x						
PGI-S	x				x		
PGI-C					x		
ICIQ-UAB 24hr	x	x	x		x	x	x

6.2.2 Concurrent PRO measures used in study

- ICIQ-MLUTS (International Consultation on Incontinence Questionnaire - Male Lower Urinary Tract Symptoms)
    - LUTS and bother in males (13 items)
- ICIQ-FLUTS (International Consultation on Incontinence Questionnaire - Female Lower Urinary Tract Symptoms)
    - LUTS and bother in females (13 items)
- ICIQ-LUTSqol (International Consultation on Incontinence Questionnaire - Lower Urinary Tract Symptoms Quality of Life)
    - Impact of LUTS symptoms on health related quality of life (22 items)
- PGI-S (Patient Global Impression of Severity)
    - Single rating of severity of their condition on a 4 point scale
- PGI-C (Patient Global Impression of Change)
    - Single rating of change of status of subject's symptoms on 7 point scale

6.2.3 Translation

The PRO measures were completed in the language of the respective country in which they were administered using linguistically validated versions (English, Dutch and German). The Dutch and German ICIQ-UAB versions were developed from UK English using linguistic

validation methodology to ensure conceptual equivalence [175]. The process was outsourced and coordinated by MAPI, an external patient-centered research company that has expertise in translation of PRO measures [174]. The author was responsible for reviewing the back translations to ensure the original concepts were retained.

6.2.4 Recruitment

Participants were recruited from sites in the UK, Netherlands and Germany using a purposive sampling approach. Study management was outsourced to a company (Pharmerit), based in the Netherlands. The author was responsible for the recruitment of subjects and administration for the Bristol site.

Patients were sent a postal invitation to participate in the study, containing a patient information letter. If interest was expressed then informed consent was taken by further postal correspondence. A case report form was also signed and completed and checked for conformity with the inclusion and exclusion criteria by the author. After written informed consent was obtained from patients, the study package containing the questionnaires was sent out to subjects to be completed.

6.2.5 Inclusion and exclusion criteria

Patients were recruited using the following inclusion and exclusion criteria.

- Male or female and at least 18 years of age.
- Institutional Review Board (IRB)-/Independent Ethics Committee (IEC)-approved written Informed Consent and privacy language as per national regulations must be obtained from the subject prior to any study-related assessments.
- Patient was willing and able to complete the PRO instruments.
- Patient had a primary diagnosis of DU for at least 3 months, with stable symptoms.
- Patient was able to void spontaneously and pass at least part of their urine by voiding.
- Patient was willing and able not to make changes to concomitant medication and life style, which may have an effect on lower urinary tract symptom (LUTS) (e.g., start urological treatment or change fluid intake) during the study period (10 days).

Patients were ineligible for this pilot assessment if they met any of the following exclusion criteria:

- Stress urinary incontinence (SUI), unless mild, and DU were the predominant conditions.
- Detrusor Overactivity (DO) unless mild and DU is the predominant condition.
- Patient had a neurological cause for DU (e.g., neurogenic bladder or systemic or central neurological disease such as multiple sclerosis and Parkinson's disease). Patients with diabetic neuropathy could be enrolled.
- Patient had received treatment that may impact the pressure flow dynamics since the last pressure flow study (PFS), which confirmed DU diagnosis.
- Patient was participating in an investigational trial.

Further criteria:

- Patients with or without a post-void residual were included in the pilot assessment, if they met all the above-mentioned inclusion and exclusion criteria.
- Patients using intermittent self-catheterization were included in the pilot assessment, if they met all the above-mentioned inclusion and exclusion criteria.
- Patients with an acontractile bladder were included in the pilot, if they met all the above-mentioned inclusion and exclusion criteria.

The same urodynamic inclusion and exclusion criteria as for the concept elicitation and cognitive interviews were used to select participants with a primary urodynamic diagnosis of DU. As before, prospective patients for the Bristol site were screened for suitability by the author.

The pressure flow study (PFS) parameters used to identify potential patients were the bladder contractility index (BCI) and bladder outlet obstruction Index (BOOI) for males, and detrusor pressure at maximum flow ($p_{det}Q_{max}$), and maximum flow rate (Q_{max}) for females. Patients were categorised in the diagnostic groups to facilitate the analyses of the data, as detailed in Table 10.

Table 16. Diagnostic group inclusion criteria.

DU (only) • Patients with a confirmed diagnosis of DU on urodynamics and a diagnosis of no other co-existing urological conditions Males: BCI <100 BOOI <20 Females: $p_{det}Q_{max}$ <20 cmH_20 Q_{max} <15ml/s
DU + mild DO • DU + co-existing mild DO (based on investigator unrodynamic assessment and patient records)
DU + Mild SUI/USI • DU + co-existing mild USI/SUI (based on investigator urodynamic assessment and patient records)
DU + BOO-Equivocal (males only) • BCI <100, BOOI ≥20–<40
DU + BOO (males only) • BCI <100, BOOI ≥40
DU + Mixed (other) • DU + BOO-E and Mild SUI/USI or mild DO as defined by the above

Abbreviations: Detrusor pressure at maximum flow ($p_{det}Q_{max}$), maximum flow rate (Q_{max}), bladder contractility index (BCI) calculated by BCI=$p_{det}Q_{max}$+5Q_{max}, Bladder Outlet Obstruction Index (BOOI) calculated by BOOI=$p_{det}Q_{max}$-2Q_{max}.

6.2.6 Sample size

A minimum sample size of 40 is advised in order to obtain reliable estimates when assessing item performance and tests of reliability in pilot studies such as the current study [206]. The aim was to recruit at least 50 participants to fulfil these criteria.

6.2.7 Scoring

A pragmatic approach to the scoring of the ICIQ-UAB was taken in order to undertake the current analyses, as scoring algorithms based on psychometric evidence will be derived at a later stage. The response option of 'never' or 'not at all' was scored 0 with subsequent options scoring 1, 2, 3 or 4 for the response options of increasing severity. For the ICIQ-UAB 1-week, items 1-3 (medical history) were not included in the calculation of the overall score, as these would not be expected to change. Item 22 (bowel associations) was not included due to the first part of the item being dichotomous (yes/no). Item 23 (clustering of symptoms) was on a nominal scale so was not included in the total score. Item 31 was an overall question on impact bother with a 0-10 Likert scale so was also not included in the score. For ICIQ-UAB

24hrs the corresponding item 19 (bowel associations) and item 20 (clustering of symptoms) were not included. Table 17 summarises how the scores were calculated for each domain.

Table 17. Scoring in each of the versions.

Domain	ICIQ-UAB 1-week		ICIQ-UAB 24hrs	
	Item number	Range of possible total scores	Item number	Range of possible total scores
Total score	(4-21)+(24-30)	0-100		
Symptom items	4-21	0-72	1-18	0-72
Impact items	24-30	0-28		

6.2.8 Psychometric methodology

The following details the analytic methodology that was performed on the data. These included descriptive analyses (e.g. missing data, floor and ceiling effects) and assessment of the instrument's psychometric properties (aspects of validity and reliability).

Missing data

The percentage of missing responses was calculated for each administration. A level of missing data of 3-5% was generally considered acceptable [149,150].

Floor and ceiling effects

Items were considered to have a floor effect or ceiling effect if more than X% of respondents answered the highest or lowest option respectively, where X is 100/the number of response options.

Bother

The bother scores associated with each item ranged from 0 – 10, where score of 0 represents no bother and a score of 10 maximum bother. The reported score was only included in the calculation of the average bother score for each item if the patient experienced the symptom/impact (i.e. did not answer 'not at all' on the item). Thus, each item has a different denominator depending on the number of patients who experienced the symptom/impact. The mean bother scores were calculated at baseline (day 1) for each item in the ICIQ-UAB 1-week and ICIQ-UAB 24 hrs (Table 20). In addition, mean scores were calculated for known

groups of severity (PGI-S score at baseline). This was to assess sensitivity of bother scores to severity of condition.

Construct validity

Although the literature links UAB with older adults [31,207,208], the link of impaired detrusor contractility with ageing *per se* has yet to be conclusively demonstrated [18,41]. However, urinary retention, hesitancy, and incontinence have been associated with UAB and advancing age [16,25]. It is also known that symptom of nocturia is highly correlated with increasing age in the general population [47,48]. The sample was stratified by participants ≤65 years of age and those >65 years, to investigate a potential positive relationship of age and reported symptom occurrence.

The prevalence of UAB symptoms in males and females is poorly understood in the literature. A recent retrospective database analysis of patients referred for pressure flow studies (PFS) by Gammie et al. [21], elucidated some of the symptom prevalence for men and women with DU compared to those with normal PFS and BOO. The difference in symptom prevalence between men and women with DU was not compared. However, stress incontinence was reported by 25% of men and 79% of women with DU in the sample. Urinary incontinence is also known to be less prevalent in the male general population [153,209,210] so may provide some proxy for comparison in the absence of data for other symptoms that are perhaps more associated with UAB. The sample was investigated for any relationship between symptom prevalence and gender.

Clean intermittent catheterisation (CIC) or intermittent self-catheterisation (ISC) has been recommended and used successfully for neuropathic bladder problems and bladder drainage for patients for a number of years [211]. However, the apparent clinical benefits to patients with UAB are not well documented, despite being the only current intervention available for UAB [25] (sacral neuromodulation (SNM) also shows promise in appropriately selected patients [19]). The use of ISC was investigated for association of symptom relief with reported symptoms.

Incomplete bladder emptying is considered a classic symptom of UAB [25,26]. The presence of a PVR was investigated for a any association with reported symptoms.

Criterion validity

There are no existing validated PRO measures to assess UAB against which the ICIQ-UAB can be compared, as is most often the case when the development of a new PRO is justified. However, concurrent validity was assessed by evaluating responses to items within the ICIQ-UAB 1-week at baseline for associations with responses to related items in other PRO measures of known validity. Items within the ICIQ-UAB were paired with similar items in the ICIQ-MLUTS, ICIQ-FLUTS and ICIQ-LUTSqol. A Spearman's correlation coefficient of >0.7 was used as the accepted cut-off to indicate a correlation [212].

It was hypothesised there should be a very good correlation between paired items which have identical shared target concepts. Reasonable correlations were expected between paired items with similar or 'best fit' measured concepts.

Known groups analysis

Known groups of severity were designated by the responses given to PGI-S data at baseline (no symptoms n=7, mid n=11, moderate n=27, severe n=4). It was hypothesised that the mean score of both versions of the ICIQ-UAB would increase with the level of severity.

Internal consistency

Cronbach's alpha was calculated for the baseline scores for all respondents (n=54) in order to explore the extent of the homogeneity of the items, or the ability of the questionnaire to measure the same concept [103]. A Cronbach's alpha of ≥0.7 was considered acceptable and >0.9 to possibly indicate redundancy within the item pool [154].

Test –retest reliability

The stability of the items was measured by the scores obtained on separate occasions in a 'stable' population, or over a period of time where there are not expected to be any change in their symptoms [99]. The test-retest responses were analysed using only the data from the patients who gave an answer 'no change' to the PGI-C on day 8. The PGI-C items have been widely used to assess the subject's perspective of improvement, and have been shown to have validity in a variety of clinical trials [213,214].

The Intraclass correlation coefficient (ICC) was calculated between scores returned on day 1 and day 8 for ICIQ-UAB 1-week, and between successive pairs of administrations for ICIQ-UAB 24 hrs. The accepted criterion for reliability was ≥0.7, and >0.9 was considered excellent [154].

The extent of agreement between item-level responses for repeat administrations was also evaluated. The kappa coefficient was used to give a chance-corrected measure of agreement at the item-level (Kappa statistic (κ): 0 = poor, 0.01-0.20 = slight, 0.21-0.40 = fair, 0.41-0.6 = moderate, 0.61-0.8 = substantial, and 0.81-1 = almost perfect agreement [157]).

6.3 Results

A total of 54 patients with a primary diagnosis of DU were recruited from 4 sites in the UK (n=29), 3 sites in the Netherlands (n=16) and 1 site in Germany (n=9). The sample was entirely Caucasian and predominantly male (80%), had a mix of employment status, were diagnosed with or without co-existing urological conditions and had a range of self-reported severity of symptoms and urological medications. The summary of demographic characteristics and urodynamic parameters is included in Table 18.

Table 18. Summary of demographic characteristics and pressure flow study parameters.

Clinical or demographic characteristic	Total sample (n=54)
DU (only) n (%)	33 (61)
DU + mild DO n (%)	10 (19)
DU + mild BOO n (%)	8 (15)
DU + mixed (other) n (%)	3 (5)
Mean age and range (years)	61.2 (19-88)
Gender – male n (%)	43 (80)
ISC n (%)	10 (19)
PVR >30ml n (%)	36 (67)
PVR <30ml n (%)	16 (30)
No PVR (0ml) n (%)	11 (20)
BVE(%) mean (SD)	64.9 (31.3)
BCI mean (males only) (SD)	76.0 (19.3)
BOOI mean (males only) (SD)	19.4 (12.1)
$P_{det}Q_{max}$ (cmH20) mean (SD)	30.5 (12.6)
Q_{max} (ml/sec) mean (SD)	8.0 (3.5)

Abbreviations: Standard Deviation (SD). Intermittent self-catheterisation (once a day or more) (ISC), Detrusor pressure at maximum flow ($p_{det}Q_{max}$), maximum flow rate (Q_{max}), bladder contractility index (BCI) calculated by BCI=$p_{det}Q_{max}$+5Q_{max}, Bladder Outlet Obstruction Index (BOOI) calculated by BOOI=$p_{det}Q_{max}$-2Q_{max}, Post Void Residual (PVR), Bladder Voiding Efficiency (BVE) calculated by voided volume/total bladder capacity) × 100. Detrusor Overactivity (DO).

6.3.1 Missing data

Missing data at baseline for the ICIQ-UAB 1-week was very low (no responses missing or <3%) for all items and administrations, with the exception of item 23 'Were your symptoms worse at particular times? Please tick those that apply' which had 7% missing data at baseline and on retest at day 8. No patients ticked multiple boxes for this item. The missing data for ICIQ-UAB 24hr was also very low for all items at all administrations (no responses missing or <5%). At day 1, one patient did not complete 6 items and another patient missed the whole day 10 administration.

6.3.2 Floor and ceiling effects

All three of the medical history items (Q1-3) in the ICIQ-UAB 1-week showed strong floor effects. Item 2, 'urinary tract infections' also showed a ceiling effect, of which 26% of responses were the 'three or more times' response option. A total of 9/20 symptom items showed a floor effect. The largest floor effects (>50%) were observed in the 'incontinence', 'temporarily unable to pass urine' and 'post-micturition dribble' items at 70%, 61% and 50% respectively. All impact items showed strong floor effects with over 40% of responses given as 'not at all' in 6/7 items. The highest response option was not used in the items for 'incontinence', 'urgency', 'nocturnal voids', 'daytime frequency' and 'temporarily unable to pass urine'. The second response option for the item 'length of time in bathroom' attracted 70% of the responses with the highest two response options also used very infrequently.

A total of 12/20 corresponding items in the ICIQ-UAB 24hrs showed a floor effect. The additional three items which had a floor effect in this version (compared to the ICIQ-UAB) were 'urgency', 'straining towards end of void' and 'small volume of urine per void'. It should be noted that these additional items were only marginally over the threshold of 26%, 20% and 22% respectively. The same three medical history items had floor effects above 50% as in the ICIQ-UAB. The highest response option was not used in the 'length of time in the bathroom' and 'temporarily unable to pass urine' items.

Table 19 shows the percentage of responses in the lowest response option, and Figure 7 and Figure 8 shows the frequency distribution of the ICIQ-UAB 1-week responses by item, for the symptom and impact items respectively. Figure 9 shows the ICIQ-UAB 24hrs responses at baseline.

Table 19. The percentage of responses in lowest response option for each version at baseline. A shading of red indicates a floor effect (≥20% of responses were in lowest category).

Item (ICIQ-UAB 1-week/ICIQ-UAB 24hrs)	% of responses in lowest response option	
	ICIQ-UAB 1-week	ICIQ-UAB 24hrs
Acute retention (Q1)	83	
UTIs over last month (Q2)	59	
Self-catheterisation (Q3)	80	
Hesitancy (Q4/Q1)	9	13
Need to concentrate to void (Q5/Q2)	28	20
Small volume of urine per void (Q6/Q3)	15	22
Post-micturition dribble (Q7/Q4)	50	52
Incontinence (Q8/Q5)	70	67
Need to immediately re-void (Q9/Q6)	28	37
Incomplete emptying (Q10/Q7)	13	19
Intermittency (Q11/Q8)	15	19
Straining to start (Q12/Q9)	24	30
Straining towards end of void (Q13/Q10)	9	20
Slow stream (Q14/Q11)	15	15
Urgency (Q15/Q12)	19	26
Nocturnal voids (Q16/Q13)	9	11
Daytime frequency (Q17/Q14)	7	7
Reduced bladder sensation (Q18/Q15)	24	20
Waiting in bathroom after voiding (Q19/Q16)	24	26
Length of time in bathroom (Q20/Q17)	15	15
Temporarily unable to pass urine (Q21/Q18)	61	67
Bowel association (Q22/Q19)*	78	82
Clustering of symptoms (Q23/Q20)**	n/a	n/a
Impacts		
Planning life around toilets (Q24)	41	
Social life (Q25)	50	
Nocturia/nocturnal voids impact (Q26)	41	
Physical activities (Q27)	46	
Feelings about self (Q28)	46	
Embarrassment (Q29)	43	
Fluid intake (Q30)	30	

*Includes patients who answered 'no' in to the initial dichotomous yes/no option.

** This item has nominal response options.

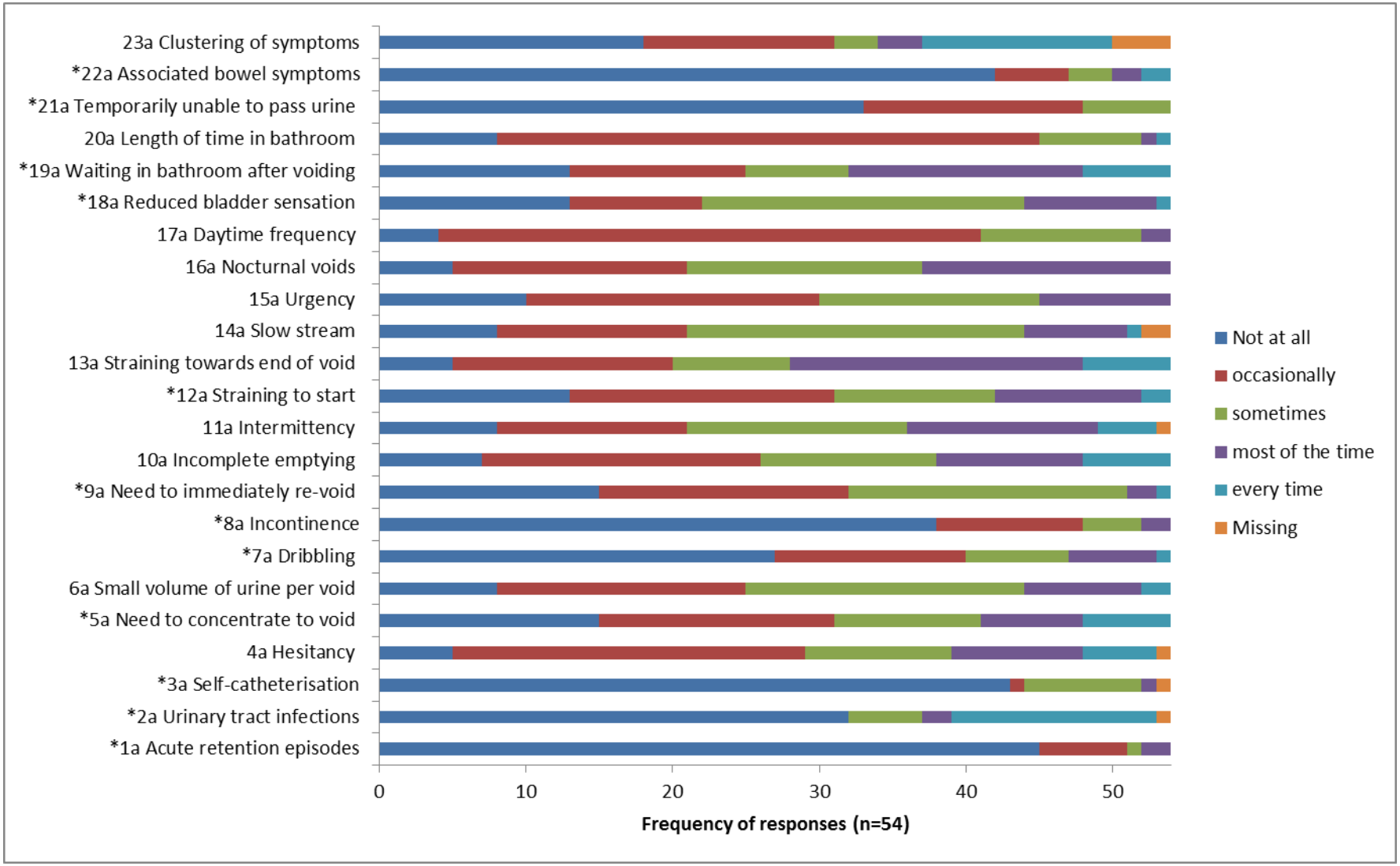


Figure 7. ICIQ-UAB 1-week symptom items showing frequency distribution of responses at baseline. *Items which met criteria for floor effect.

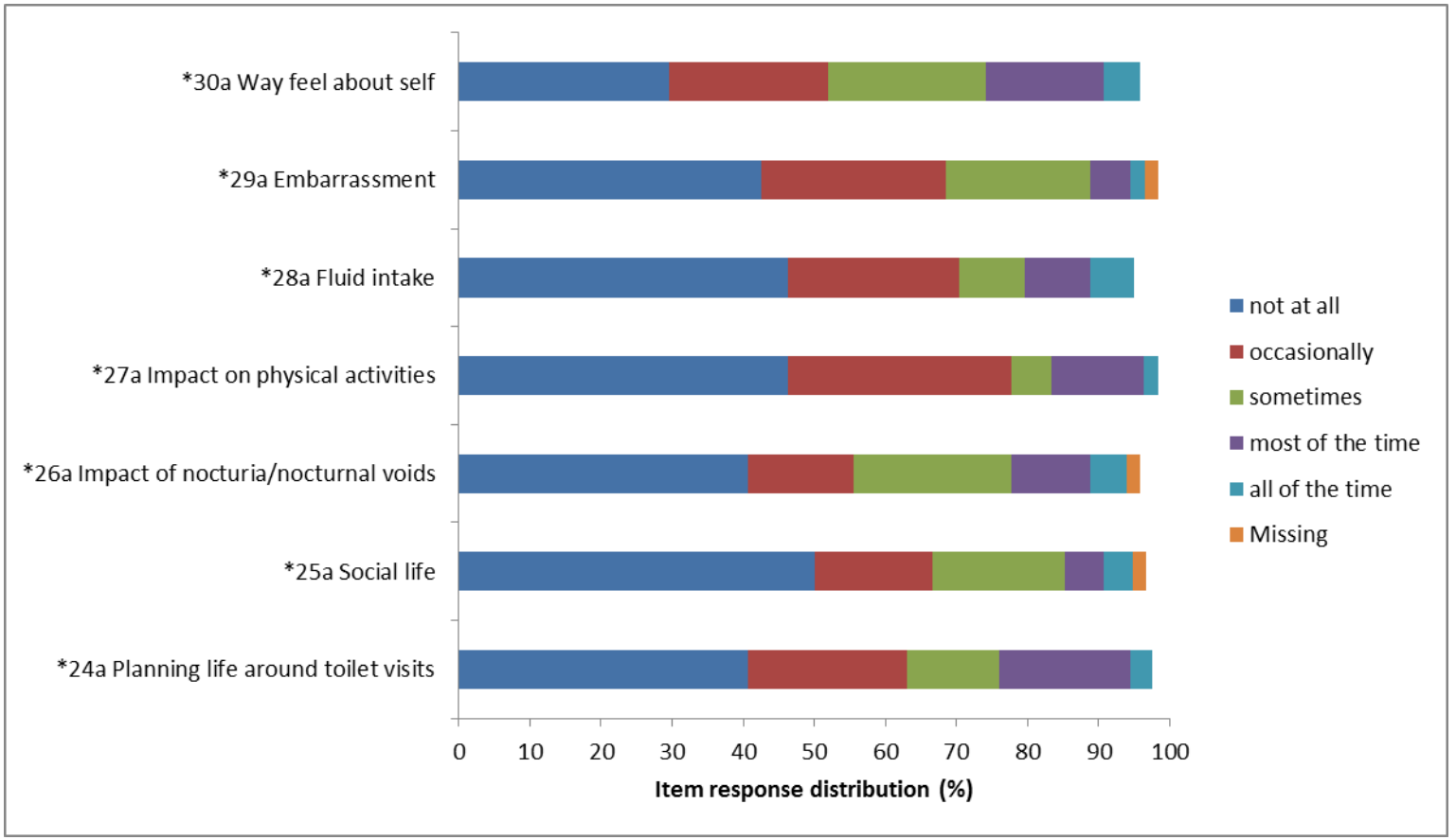


Figure 8. ICIQ-UAB impact items showing frequency distribution of responses at baseline.

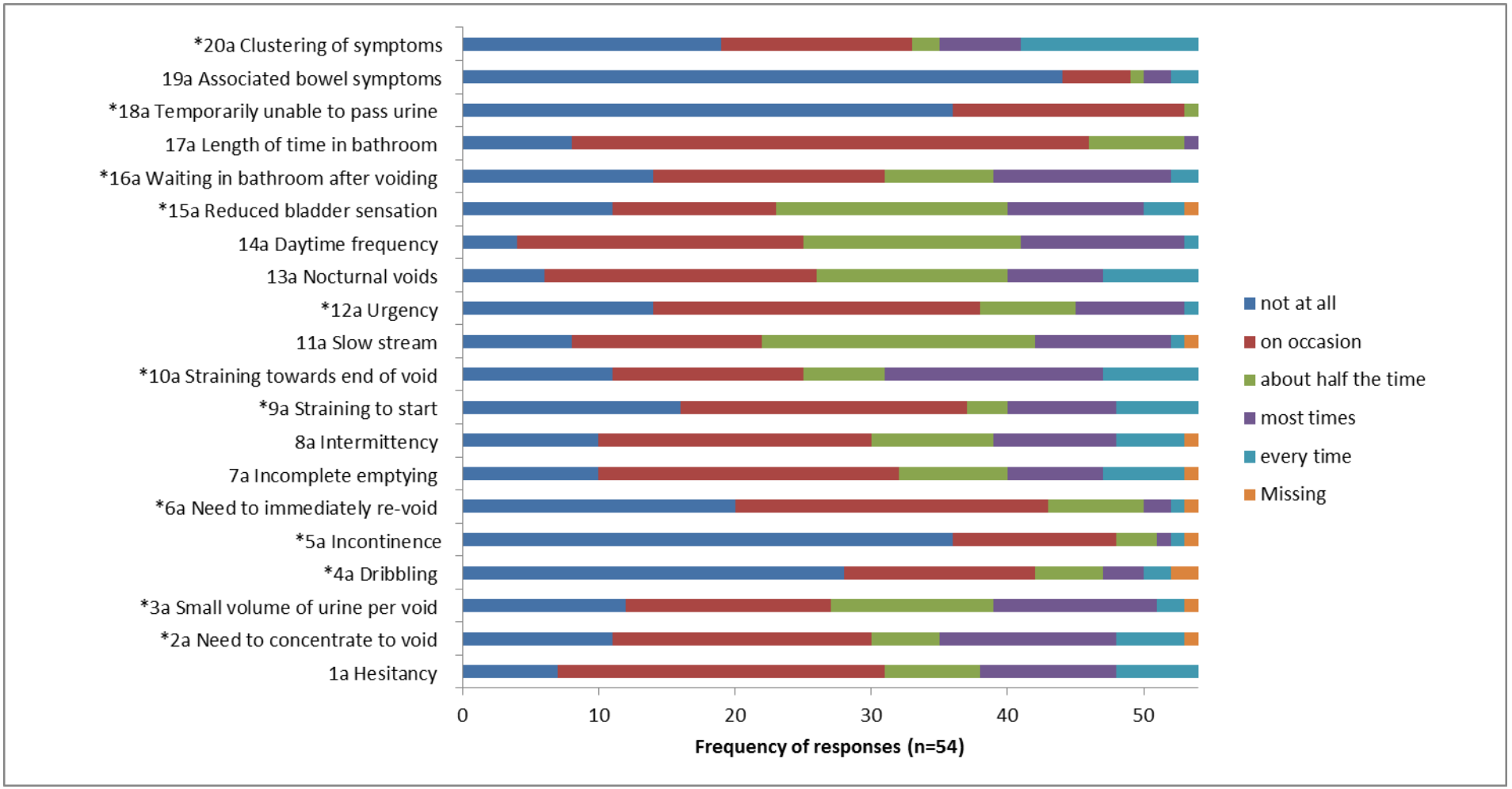


Figure 9. ICIQ-UAB 24hrs items showing distribution of responses at baseline. *items which met criteria for floor effect.

6.3.3 Bother

The medical history items (Q1-3) showed relatively high bother ranging from 6.4-7.1. Symptom mean scores for the ICIQ-UAB 1-week ranged from 4.5-6.9, 'clustering of worst symptoms' having the lowest mean bother score and 'incontinence' the highest. The ICIQ-UAB 24hrs mean scores ranged from 4.5-6.8, 'clustering of worst symptoms' having the lowest mean bother score again and 'bowel associations' with the highest. The mean bother scores for corresponding items in each of the recall period versions were very similar. Figure 10 shows the mean bother scores were lower in the patients who rated their symptoms as none/mild on the PGI-S at baseline than the patients who rated their symptoms as moderate/severe.

Table 20. Mean bother scores at baseline day 1 (for those who experienced the symptom) for ICIQ-UAB and ICIQ-UAB 24hrs.

Item (ICIQ-UAB 1-week/ICIQ-UAB 24hrs)	ICIQ-UAB 1-week	ICIQ-UAB 24hrs
	Mean score (SD) n=54	Mean score (SD) n=54
Acute retention (Q1)	6.4 (3.4)	
UTIs over last month (Q2)	7.7 (2.5)	
Self-catheterisation (Q3)	7.1 (2.3)	
Hesitancy (Q4/Q1)	4.6 (2.8)	5.0 (3.1)
Need to concentrate to void (Q5/Q2)	5.2 (2.7)	5.3 (3.0)
Small volume of urine per void (Q6/Q3)	5.6 (2.7)	5.8 (3.0)
Post-micturition dribble (Q7/Q4)	6.6 (2.4)	6.6 (2.5)
Incontinence (Q8/Q5)	6.9 (2.6)	6.3 (2.7)
Need to immediately re-void (Q9/Q6)	6.3 (2.3)	6.6 (2.5)
Incomplete emptying (Q10/Q7)	5.8 (2.8)	6.0 (3.0)
Intermittency (Q11/Q8)	5.7 (2.6)	5.4 (3.0)
Straining to start (Q12/Q9)	5.8 (2.8)	5.7 (3.2)
Straining towards end of void (Q13/Q10)	5.8 (2.7)	6.3 (2.9)
Slow stream (Q14/Q11)	5.5 (2.8)	5.2 (3.0)
Urgency (Q15/Q12)	6.1 (2.6)	6.3 (2.4)
Nocturnal voids (Q16/Q13)	5.8 (3.0)	5.5 (3.1)
Daytime frequency (Q17/Q14)	4.9 (3.1)	4.9 (3.3)
Reduced bladder sensation (Q18/Q15)	4.9 (3.1)	5.1 (3.0)
Waiting in bathroom after voiding (Q19/Q16)	5.6 (2.8)	5.5 (2.6)
Length of time in bathroom (Q20/Q17)	5.0 (2.9)	5.1 (3.3)
Temporarily unable to pass urine (Q21/Q18)	6.0 (3.2)	5.8 (3.1)
Bowel association (Q22/Q19)	6.5 (2.7)	6.8 (3.0)
Clustering of symptoms (Q23/Q20)	4.5 (3.6)	4.5 (3.7)
Planning life around toilets (Q24)	7.1 (2.2)	
Social life (Q25)	6.4 (2.5)	
Nocturia/nocturnal voids impact (Q26)	6.7 (2.5)	
Physical activities (Q27)	6.1 (2.5)	
Feelings about self (Q28)	6.3 (3.1)	
Embarrassment (Q29)	6.6 (2.7)	
Fluid intake (Q30)	5.8 (2.6)	
Overall impact (Q31)	5.1 (3.3)	

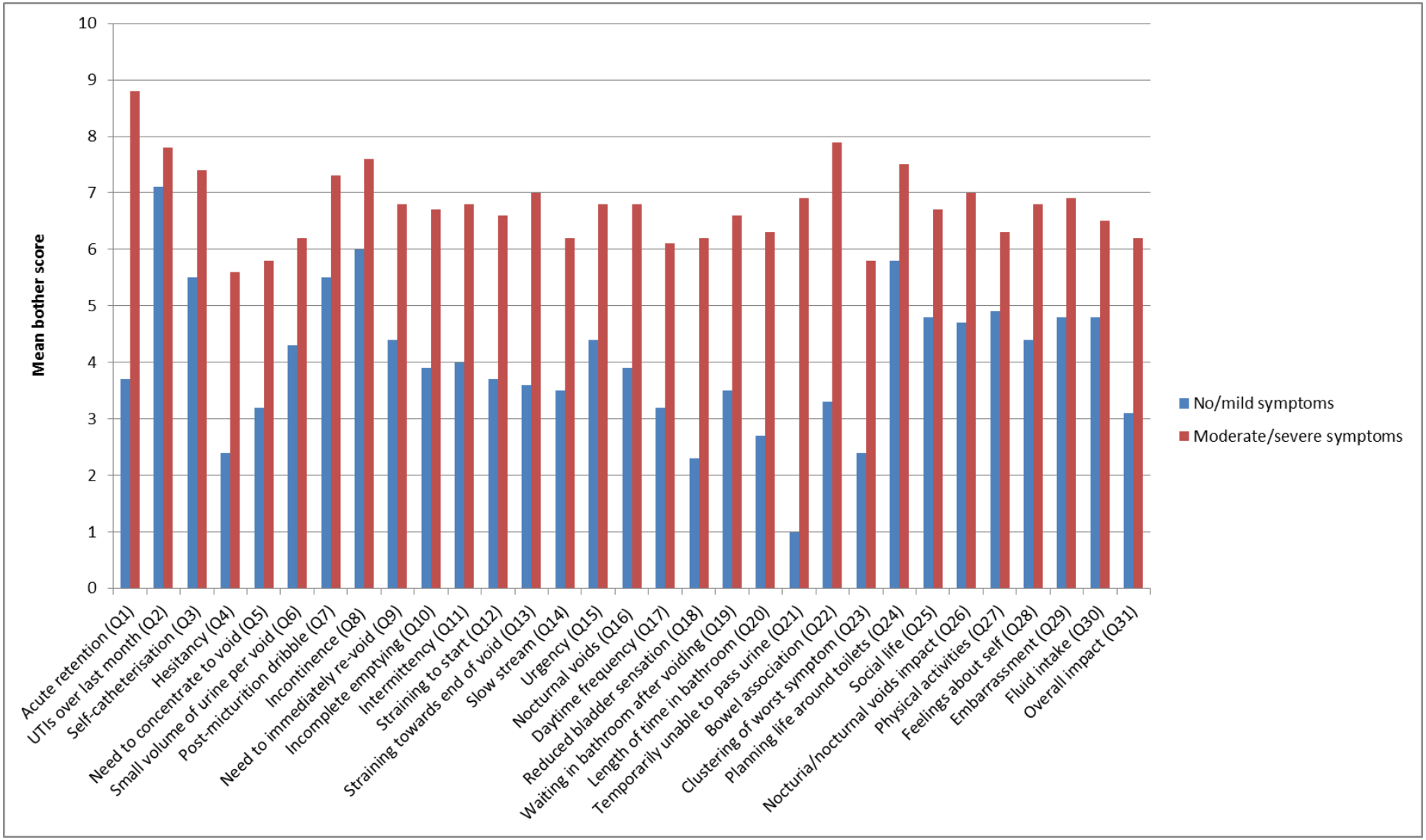


Figure 10. Mean bother scores for ICIQ-UAB 1-week stratified by those with no/mild symptoms and those with moderate/severe symptoms as determined by the PGI-S at baseline.

6.3.4 Construct validity

Table 21 and Table 22 show the percentage of patients reporting symptoms and impacts at baseline (in the ICIQ-UAB 1-week), by the investigated subgroups: age, gender, self-catheterisation and post void residual. The pairs of values highlighted in green have a proportionate significant difference ($p<0.05$). Unless indicated, the percentage of patients who answered 'sometimes', 'most of the time' or 'every time' are given.

The only symptom/sign that was correlated by age was 'temporarily unable to pass urine', which was less prevalent in older participants.

Females were found to be more likely to report increased daytime frequency and incontinence. The item for 'bowel associations' was reported by a higher percentage of males.

Those who practice ISC were found to have significantly fewer reported nocturnal voids. This did not quite reach statistical significance for the reported impact of nocturnal voids (Table 22). In addition, those who practised ISC were more likely to report historical urinary tract infections.

No significant difference was found between reported symptoms in participants with a urodynamically confirmed PVR. This included the sensation of incomplete emptying.

There were no significant differences in reported impacts in any of the sub-groups, except for men who showed a higher tendency to report a greater impact on the way they felt about themselves.

Table 21. The percentage of patients reporting symptoms on the ICIQ-UAB 1-week at baseline in each sub-group by age, gender, self-catheterisation and post void residual. Highlighted in green are the pairs which have a significant proportionate difference (p<0.05).

Symptom	% patients reporting symptom							
	Age ≤65 years (n=28)	Age >65 years (n=26)	Male (n=43)	Female (n=11)	No ISC (n=44)	ISC (n=10)	PVR <30ml (n=16)	PVR (n=36)
Acute retention episodes (Q1)	18	15	21	0	16	20	25	14
Urinary tract infections (Q2)	43	35	33	64	30	80	31	39
Self-catheterisation (Q3)	18	19	12	45			6	22
Hesitancy (Q4)	54	35	42	55	41	60	56	36
Need to concentrate to void (Q5)	54	31	44	36	39	60	50	36
Small volume of urine per void (Q6)	54	54	51	64	48	80	56	50
Dribbling (Q7)	25	27	26	27	27	20	25	25
Incontinence (Q8) (Occasionally or more)	25	15	23	55	30	30	13	36
Need to immediately re-void (Q9)	43	38	42	36	39	50	38	42
Incomplete emptying (Q10)	46	62	51	55	50	60	56	47
Intermittency (Q11)	61	58	53	82	61	50	75	50
Straining to start (Q12)	54	31	42	45	43	40	44	39
Straining towards end of void (Q13)	68	58	58	82	52	70	69	58
Slow stream (Q14) (reduced or more)	57	58	58	55	57	60	69	53
Urgency (Q15)	46	42	44	45	43	50	44	44
Nocturnal voids (Q16) (twice or more)	54	69	60	64	70	20	63	61
Daytime frequency (Q17) (10-12 times or more)	21	27	16	55	20	40	19	28
Reduced bladder sensation (Q18) (difficult or more)	64	54	53	82	55	80	56	58
Waiting in bathroom after voiding (Q19)	57	50	51	64	52	60	69	44
Length of time in bathroom (Q20) (6-10 minutes or more)	21	12	16	18	18	10	6	19
Temporarily unable to pass urine (Q21) (occasionally or more)	54	23	40	36	36	50	38	39
Associated bowel symptoms (Q22)	32	38	42	9	36	30	44	33
Clustering of symptoms (Q23)	57	62	65	36	64	40	69	56

Table 22. The percentage of patients reporting impact on the ICIQ-UAB 1-week at baseline in each sub-group by age, gender, self-catheterisation and post void residual. Highlighted in green are pairs which have a significant proportionate difference (p<0.05).

Impact	Patients reporting impact (%)							
	Age ≤65 years (n=28)	Age >65 years (n=26)	Male (n=43)	Female (n=11)	No ISC (n=44)	ISC (n=10)	No PVR (n=16)	PVR (n=36)
Planning life around toilets (Q24)	64	54	65	55	59	60	63	56
Social life (Q25)	43	54	49	45	45	60	63	39
Impact of nocturia/ and/or nocturnal voids (Q26)	50	65	60	45	64	30	75	50
Impact on physical activities (Q27)	54	58	53	55	52	60	44	56
Fluid intake (Q28)	61	46	56	27	52	60	63	47
Embarrassment (Q29)	57	54	58	45	55	60	63	50
Way feel about self (Q30)	75	65	79	36	73	60	75	67

6.3.5 Criterion (concurrent) validity

The ICIQ-UAB 1-week symptom items relating to hesitancy, need to concentrate to void, straining to start or towards end, slow flow, intermittency, sensation of incomplete emptying, post-micturition dribble, nocturnal voids, urgency and daytime frequency all showed a good correlation of >0.7 with ICIQ-MLUTS and/or ICIQ-FLUTS corresponding items (Table 23 and Table 24). The ICIQ-UAB 1-week 'incontinence' item was most correlated with the ICIQ-MLUTS 'incontinence when cough or sneeze' item. Items which showed poor or moderate correlations (<0.7) all were 'best fit' comparisons which did not measure identical concepts (Table 23).

Table 23. Correlations of matched ICIQ-UAB 1-week items with ICIQ-MLUTS items.

ICIQ-MLUTS item description	**ICIQ-UAB 1-week item description**	**Spearman's correlation coefficient**
Hesitancy (Q1)	Hesitancy (Q4)	0.81
	Need to concentrate to void (Q5)	0.72
Straining (Q2)	Straining to start (Q12)	0.69
	Straining towards end (Q13)	0.70
Slow stream (Q3)	Slow stream (Q14)	0.72
Intermittency (Q4)	Intermittency (Q11)	0.71
Incomplete emptying (Q5)	Sensation of incomplete emptying (Q10)	0.71
	Waiting in bathroom after voiding (Q19)	0.50
Urgency (Q6)	Urgency (Q15)	0.67
Incontinence before getting to the toilet (Q7)	Incontinence (Q8)	0.38
Incontinence when cough or sneeze (Q8)	Incontinence (Q8)	0.67
Incontinence for no obvious reason (Q9)	Incontinence (Q8)	0.31
Postmicturition dribble (Q11)	Postmicturition dribble (Q7)	0.81

Table 24. Correlations of matched ICIQ-UAB 1-week items with ICIQ-FLUTS items.

ICIQ-FLUTS item description	ICIQ-UAB item description	Spearman's correlation coefficient
Nocturnal voids (Q1)	Nocturnal voids (Q16)	0.99
Urgency (Q2)	Urgency (Q15)	0.846
Daytime frequency (Q4)	Daytime frequency (Q17)	0.95
Hesitancy (Q5)	Hesitancy (Q4)	0.82
	Need to concentrate to void (Q5)	0.79
Straining (Q6)	Straining to start (Q12)	0.91
	Straining towards end (Q13)	0.70
Intermittency (Q7)	Intermittency (Q11)	0.70
Incontinence before getting to the toilet (Q8)	Incontinence (Q8)	0.62
Incontinence when active or cough or sneeze (Q10)	Incontinence (Q8)	0.50
Incontinence for no obvious reason (Q11)	Incontinence (Q8)	0.417

The ICIQ-UAB 1-week impact items all showed reasonable or very good correlations of >0.6 with the matched ICIQ-LUTSqol items, except for 'impact of nocturia and/or nocturnal voids' (Table 25).

Table 25. Correlations of matched ICIQ-UAB 1-week items with ICIQ-LUTSqol items.

ICIQ-LUTSqol item description	ICIQ-UAB item description	Spearman's correlation coefficient
Impact on physical activities (Q5)	Impact on physical activities (Q27)	0.70
Impact on ability to travel (Q6)	Planning life around toilet visits (Q24)	0.60
Impact on social life (Q7)	Impact on social life (Q25)	0.82
Impact on sleep (Q15)	Impact of nocturia and/or nocturnal voids (Q26)	0.377
Fluid intake (Q18)	Fluid intake (Q30)	0.74
Embarrassment (Q21)	Embarrassment (Q29)	0.68
Feel bad about yourself (Q14)	Way feel about self (Q28)	0.61
Overall impact (Q22)	Overall impact (Q31)	0.83

6.3.6 Known groups analysis

The mean score increased with the level of severity in both versions. There was a clear differentiation of mean scores between 'no/mild symptoms' and 'moderate/severe symptoms'. The analysis of variance showed a significant difference between the mean scores of the severity groups in all domains ($p < 0.05$) for both recall period versions. The impact score contained a clear outlier which was removed for the purposes of the analysis (patient 305 responded on the PGI-S as mild symptoms but had high total impact score of 20). Table 26 and Table 27 show the results for each of the two versions tested.

Table 26. Mean score of ICIQ-UAB 1-week domains when stratified by known group as determined by the PGI-S at baseline.

ICIQ-UAB 1-week domains	Known Groups	Mean score	SD	*P*-value
Total score	No-Symptom	19.7	13.6	0.003
	Mild	25.5	11.7	
	Moderate	40.0	13.3	
	Severe	40.3	5.7	
Symptom items	No-Symptom	19.6	8.78	0.004
	Mild	19.8	8.67	
	Moderate	30.0	9.00	
	Severe	29.3	6.40	
Impact items	No-Symptom	1.33	1.21	0.003 (with outlier removed)
	Mild	5.64	4.86	
	Moderate	9.56	5.67	
	Severe	11	4.97	

Table 27. Mean score of ICIQ-UAB 24 hrs when stratified by known group as determined by the PGI-S at baseline.

ICIQ-UAB 24hr	Known Groups	Mean score	SD	*P*-value
ICIQ-UAB 24hr	No-Symptom	15.3	8.38	0.001
	Mild	19	9.62	
	Moderate	30.5	10.4	
	Severe	31.8	7.76	

6.3.7 Internal consistency

As Table 28 shows, all scores in both versions were within the accepted range of reliability Cronbach's alpha coefficient ≥0.7.

Table 28. Internal consistency for each version.

	Cronbach's alpha coefficient	
Domain	**ICIQ-UAB 1-week**	**ICIQ-UAB 24hrs**
Total score	0.88	
Symptom	0.84	0.87
Impact	0.82	

6.3.8 Test-retest reliability

A total of 12 out of the 54 subjects reported that there had been an alteration in their symptom severity according to the PGI-C at day 8. These patients were excluded from the analysis so the following results represent the test-retest responses of a pool of 42 'stable' participants, as determined by the PGI-C.

Test-retest scores and domains

Scores were reproducible for the ICIQ-UAB 1-week over the test-retest period with a mean score difference of 1.7 for the total score on the total 100-point scale, and 1.4 on the symptom item scale (0-72 point scale) (Table 29). The impact items mean score difference was identical (0-28 point scale). The largest mean difference between consecutive days for the ICIQ-UAB 24hrs was 2.6 (0-72 point scale) between day 1 and day 2 (Table 30).

Both versions returned ICCs of >0.85 in domains and pairs of administrations, indicating reliability which was comfortably above the acceptable threshold of 0.7 (Table 29 and Table 30). Repeated administrations of the ICIQ-UAB 24hrs resulted in a higher ICC, the final pair of administrations (day 9 vs day 10) returning the highest ICC of 0.94 (Table 30).

Table 29. Reliability of the ICIQ-UAB 1-week scores.

ICIQ-UAB 1-week	Day 1 (mean score)	Day 8 (mean score)	ICC (n=42)	95% CI
Total score	37	35.3	0.90	0.83-0.95
Symptom items	28.2	26.6	0.89	0.89-0.94
Impact items	8.7	8.7	0.89	0.80-0.94

Table 30. Reliability of the ICIQ-UAB 24hrs scores.

ICIQ-UAB 24hrs	Administration (mean score)						ICC (95% CI)				
	Day 1	Day 2	Day 3	Day 8	Day 9	Day 10	Day 1 vs day 8 (n=42)	Day 1 vs day 2 (n=42)	Day 2 vs day3 (n=42)	Day 8 vs day 9 (n=42)	Day 9 vs day 10 (n=41)
Total score	26.0	23.4	23.7	24.5	23.7	23.0	0.86 (0.75-0.92)	0.87 (0.75-0.93)	0.92 (0.85-0.96)	0.93 (0.87-0.96)	0.94 (0.88-0.97)

Item level agreement between paired responses

Identical agreement was not found in any of the items in either version. The three medical history items (Q1-3) showed very high agreement ≥85%. A total of 12/20 of the symptom items in both the ICIQ-UAB 1-week (Q4-23) and ICIQ-UAB 24hrs showed identical agreement in ≥60% of cases. Only 3/20 symptom items in both the ICIQ-UAB and the ICIQ-UAB 24hrs had levels of agreement of less than 50%. The items 'need to concentrate to void', 'small volume of urine per void', 'daytime frequency', 'intermittency' and 'reduced bladder sensation' showed the least agreement at 40-49% in one, or both versions. The impact items in ICIQ-UAB all showed identical agreement between 50-70%.

The medical history items were the only items to show almost perfect agreement (κ=0.81-1). A total of 4/20 symptom items in the ICIQ-UAB 1-week and 5/20 items in the ICIQ-UAB 24hrs showed substantial agreement (κ=0.61-0.8). In both versions, 10/20 of the symptom items showed moderate agreement (κ=0.41-0.6). A total of 23/31 items in the ICIQ-UAB 1-week and 15/20 items showed moderate agreement or better (κ≤0.41). The items 'need to concentrate to void', 'small volume of urine per void', 'need to immediately re-void', 'urgency', 'daytime

frequency', 'waiting in bathroom after voiding', 'reduced bladder sensation', and 'intermittency' showed fair agreement (κ=0.21-0.40) in one, or both versions. 'Incontinence' was the only item to show slight agreement, in the ICIQ-UAB 24hr (κ=0.21-0.40).

In general, most of the items in each version returned very similar statistics for test-retest. Some exceptions were the item daytime frequency which showed a clearly higher agreement and kappa statistic in the 24 hour version. The item 'intermittency' showed higher agreement and weighted kappa statistic in the 1 week version. Table 31 summarises these results.

Table 31. Test-retest analysis of each item between day 1 and day 8 for the ICIQ-UAB 1-week, and between day 1 and day 2 for the ICIQ-UAB 24hrs (Identical agreement and weighted Kappa).

Item	ICIQ-UAB 1-week		ICIQ-UAB 24hrs	
	Identical agreement n (%)	Weighted kappa (κ)	Identical agreement n (%)	Weighted kappa (κ)
Retention (Q1)	41 (98)	0.92		
Urinary Infections (Q2)	40 (95)	0.92		
Self-catheterisation (Q3)	39 (95)	0.85		
Hesitancy (Q4/Q1)	26 (63)	0.50	25 (60)	0.46
Need to concentrate to void (Q5/Q2)	20 (48)	0.34	24 (57)	0.47
Small volume of urine per void (Q6/Q3)	20 (48)	0.31	21 (50)	0.38
Post-micturition dribble (Q7/Q4)	30 (71)	0.57	28 (67)	0.48
Incontinence (Q8/Q5) (Occasionally or more)	33 (81)	0.57	27 (64)	0.20
Need to immediately re-void (Q9/Q6)	22 (52)	0.34	29 (69)	0.54
Incomplete emptying (Q10/Q7)	25 (61)	0.49	28 (67)	0.58
Intermittency (Q11/Q8)	28 (70)	0.62	17 (41)	0.23
Straining to start (Q12/Q9)	24 (57)	0.45	23 (55)	0.40
Straining towards end of void (Q13/Q10)	27 (64)	0.55	24 (57)	0.46
Slow stream (Q14/Q11)	26 (65)	0.53	30 (71)	0.63
Urgency (Q15/Q12)	22 (52)	0.34	18 (43)	0.23
Nocturnal voids (Q16/Q13) (twice or more)	30 (71)	0.63	31 (74)	0.66
Daytime frequency (Q17/Q14)	18 (43)	0.30	29 (69)	0.56
Reduced bladder sensation (Q18/Q15)	27 (64)	0.52	20 (48)	0.32
Waiting in bathroom after voiding (Q19/Q16)	22 (52)	0.39	24 (57)	0.43
Length of time in bathroom (Q20/Q17)	31 (74)	0.50	33 (79)	0.62
Temporarily unable to pass urine (Q21/Q18)	35 (83)	0.67	35 (83)	0.62
Bowel association (Q22/Q19)	6 (54)	0.42	9 (89)	0.76
Clustering of worst symptom (Q23/Q20)	26 (74)	0.64	28 (67)	0.55
Planning life around toilet visits (Q24)	27 (64)	0.50		
Social life (Q25)	27 (66)	0.50		
Impact of nocturia and/or nocturnal voids (Q26)	25 (63)	0.51		
Impact on physical activities (Q27)	29 (69)	0.54		
Fluid intake (Q28)	28 (67)	0.54		
Embarrassment (Q29)	23 (56)	0.35		
Way feel about self (Q30)	22 (52)	0.38		

6.4 Discussion

The current study assessed the psychometric properties of the ICIQ-UAB 1-week and ICIQ-UAB 24hrs in the target population. The following discusses the results in terms of the conclusions which can be made concerning the validity and reliability of the instrument, including possible modification to its content and choice of recall period. The generalisability, representativeness and limitations of the study are also discussed.

6.4.1 Validity

Missing data

Low levels of missing data levels suggest the instrument was easy to complete and well understood in both versions. In the ICIQ-UAB, item 23 performed the worst (7% missing responses) and there was a misunderstanding of the 'please tick all that apply' instruction. This item is to be considered for revision or removal.

Floor effects

Many of the items in both versions showed floor effects. This may suggest items are measuring a concept which has limited relevance to the targeted population, presents relatively rarely in the population, or only occurs in patients with relatively severe symptoms. An item which displays a strong floor effect is likely to affect the potential for sensitivity to change of the final instrument so it may be necessary to revise or consider removal [99]. Items with the largest floor effects in both versions were those relating to 'incontinence', 'temporarily unable to pass urine' and 'post-micturition dribble' in both versions. Revision of the response options for the items relating to 'incontinence', 'urgency', nocturnal voids, 'daytime frequency', 'temporarily unable to pass urine' and 'length of time in the bathroom' are to be considered, as the highest response option was not used in one, or both versions. The first response option for the item 'length of time in the bathroom' is also to be considered for revision ('less than a minute' was used infrequently so may be too short a time to be considered non-symptomatic or 'normal'). All impact items in the ICIQ-UAB may be considered for revision due to large floor effects. However, this could also be an accurate reflection of true impact of these symptoms in the studied population.

Bother score

The bother scores associated with each item were shown to be an effective measure of the 'bother' relating to each symptom, sign or impact. There was a clear relationship between high bother scores and reported severity of condition, which supports the content validity for this aspect of the instrument.

Construct validity

There was some evidence that the high proportion of responses by women for the items 'daytime frequency' and 'incontinence' in the current study were in congruence with the known higher prevalence of these symptoms in women in the general population [153,209,210]. The item 'temporarily unable to pass urine' was also related to increasing age. This is encouraging evidence that that the PRO measure is able to detect trends that is evidenced by the literature. However, correlations may be due to the purposive sample selection (all were recruited as referred symptomatic patients with UAB), and therefore may not reflect the prevalence of UAB within the background population. The high proportion of females who self-catheterised and had associated bowel symptoms may also be due to the purposive nature of the sample, or the relatively small sample size of the grouped data. The use of ISC appeared to relieve the symptom of nocturia and was tentatively as expected according to the literature [25], although this did not translate to a clear effect on the reported impact of symptoms. Those who practised ISC were also more likely to report historical urinary tract infections, for which there is a documented (but not UAB specific) association [215]. No difference was found in the reported symptoms between those with or without a PVR. There was a gender difference found for the way men felt about themselves compared to women. One reason could be that men are known to have an increased awareness of their urinary problems when urinating alongside others in urinals [8].

In summary, it was not possible to provide definitive evidence of construct validity, primarily due to the current paucity of the UAB specific literature, resulting in a lack of sufficiently well evidenced theoretical constructs by which a comparison may be made. Future epidemiological studies which elucidate the prevalence of symptoms and trends associated with DU will allow construct validity to be assessed. However, it is encouraging that the associations found were in the right direction, and the PRO measure was able to detect these evidenced trends.

Criterion (concurrent) validity

An accepted 'yard-stick' measure does not exist for UAB against which criterion validity may be evaluated. However, correlations with instruments of well-evidenced validity (ICIQ-MLUTS, ICIQ-FLUTS, ICIQ-LUTSqol) provided a quasi 'gold-standard' for the evaluation of several of the items relating to LUTS and quality of life.

The concurrent validity of selected items within the ICIQ-UAB 1-week was conclusively demonstrated with similar concepts in other instruments of known validity. The ICIQ-UAB 1-week 'incontinence' item' showed the highest correlation with the ICIQ-MLUTS 'incontinence when cough or sneeze' item. This may reflect the predominant type of incontinence experienced by the UAB sample. The few items which were moderately correlated (<0.7) with their matched counterparts in other instruments were paired with 'best fit' items, so performed as expected. For example, the ICIQ-UAB item relating to 'nocturia/nocturnal voids' asked 'how often did you feel getting up at night to urinate affected your day to day life?' whereas the ICIQ-LUTSqol was 'does your urinary problem affect your sleep?'.

Known groups

Analysis of variance between known groups of symptom severity, as determined by the PGI-S at baseline, showed a relationship between a higher level of reported symptom severity and increasing score total. Thus, the instrument was shown to be sensitive to the reported overall severity of condition.

6.4.2 Reliability

Two methods of reliability were used to evaluate the consistency of the ICIQ-UAB, internal consistency and test-retest reliability. These methods provided evidence that the score derived from the instrument is consistent with itself and that a high proportion of variability within repeated administrations of the instrument is due to the underlying scale [103].

Internal consistency

The relationship between items showed excellent reliability over the accepted range (Cronbach's alpha ≥0.7) [154] in the tested domain scales of both versions. Internal consistency of the ICIQ-UAB was towards the upper end of the accepted range (Cronbach's alpha of 0.88) so there was likely to be some redundancy within the overall instrument. This to be expected,

due to the current exhaustive inclusivity of items and supports the capacity for the future reduction in the number of items.

Test-retest reliability

Test-retest reliability in stable patients was shown to be very good with ICCs over 0.85 for both versions. This provides evidence that the instrument was capable of reliably measuring what it intends to measure, and that findings were not just due to systematic errors [103,104]. Repeated administrations of the 24 hour version resulted in greater reliability, possibly suggesting a learning effect.

Item-level reliability

The assessment of reliability at the item level allowed the evaluation of individual item test-retest performance. Most items in both versions performed well, with ≥60% of responses being identical. The items 'need to concentrate to void', 'small volume of urine per void', reduced bladder sensation and 'daytime frequency' may be considered for revision, following relatively low identical agreement performance (40-49%). The weighted kappa statistic provided evidence of 'moderate' agreement or better (κ≤0.41) in 23/31 items in the ICIQ-UAB 1-week and 15/20 in the ICIQ-UAB 24hrs. It is recognised that day to day variability of symptoms and impacts (as evidenced in the concept elicitation phase) may affect the consistency of identical responses expected, when measured on two separate occasions. However, items which performed poorly may be flagged for revision or removal from the final instrument, following further testing.

6.4.3 Scoring

The format and content of item 22 (associated bowel symptoms) and item 23 (clustering of symptoms) did not lend themselves well to scoring so it was necessary to exclude these items from contributing to the total score calculations. The first response of the item relating to 'daytime frequency' was '1-3 times' which may be considered symptomatic due to a low frequency. The scoring here is to be considered for revision to allocate this option a score of '1'. Thus, the second response option of '4-6' times is scored '0' to approximate what may be considered a 'normal' number of urinations per day.

6.4.4 Recall period

Both the 1-week and 24hr versions were very similar in terms of the psychometric properties they exhibited. Three more items in the ICIQ-UAB 24hrs had floor effects that were marginally over the threshold of 20%. However, the highest response option was not used in three more of the items in ICIQ-UAB 1-week. The average symptom domain score of the ICIQ-UAB was marginally higher than the 24 hour version, suggesting more capacity for sensitivity to change. However, the range of mean corresponding scores in the known groups analysis was slightly greater in the 24 hour version, suggesting this was more sensitive to reported severity of condition. The mean bother scores and internal consistency was very similar for both versions. In general, the test-retest identical agreement between repeated administrations was very similar, along with the chance corrected weighted kappa statistic. However, there was some evidence that a few items could be more suitable for specific recall periods. For example, the item relating to 'daytime frequency' was more reliable with a 24 hour recall period whereas the item 'intermittency' was more reliable with the 1 week recall period.

Therefore, there was no strong evidence to suggest an overall advantage of a 24 hour or one week recall period, supporting the conclusion that both versions should go on to further testing.

6.4.5 Strengths and limitations

The sample was primarily purposive in nature, that is, patients were approached that were referred for urodynamic studies at the included sites and met the inclusion/exclusion criteria. Although the epidemiology of UAB is not well understood [14,25,26], it is acknowledged the demographic characteristics are unlikely to be representative of the prevalence of UAB in the background population (by gender, age, ethnicity). For example, there was a larger proportion of males recruited (80%) which may reflect that the sample was recruited from patients who had undergone urodynamic investigation. There is a necessity for diagnostic pressure flow studies in men in order to differentiate DU patients from those with predominant BOO, in order to avoid further referral to corrective surgery for LUTS (for which there may be little benefit for DU patients) [203,204]. Women who present with these symptoms may be more likely to be referred to self-catheterisation management techniques rather than further urodynamic investigation. Nevertheless, effort was made to ensure the sample was as representative as possible of the referred patient population and of the target population in which the questionnaire is intended to measure. Although all subjects were Caucasian, a strength of the

pilot study component was that the sites were located in three different European countries, which improves the generalisability of the instrument to wider populations.

It should be acknowledged that the response options of the two versions were worded slightly differently to match the recall period set. However, these were not sufficiently different to make comparison of the two recall period versions impossible for the corresponding symptom items. Each version was independently analysed so the conclusions regarding reliability and validity remain unchanged.

There is no definitive standard for sample size in questionnaire development studies [103] and the required minimum sample size of 40 was met [206]. However, a larger sample size would have further improved the generalizability or strength of the conclusions. Nevertheless, the sample was large enough to provide adequate supportive evidence of the validity and reliability of the ICIQ-UAB at this stage of initial validation.

6.5 Instrument modification following the pilot study

A number of decisions were made surrounding items within the ICIQ-UAB, following the evaluation of the pilot study data. These changes were investigated to be understood and interpreted as intended by patients by seven additional cognitive interviews. The details of these interviews and rationale for each of the changes are given in Appendix 7.

Modifications to the questionnaire as a result of the psychometric testing were deliberately conservative. It is anticipated that the number of items will be reduced using PRO data from larger clinical trials, as well as the evaluation of responsiveness to change and item scoring. The decisions and main changes made are summarised as follows:

- The 24 hour recall period was retained for the symptoms items for future clinical trials. A longer recall period of 1 week was retained for the impact items due to patient preference. However, both versions will remain available for further testing.
- Corresponding response options were selected from the 24 hour recall period version.
- Minor revisions to wording were made to aid the clarity of interpretation and address floor effects in the following item stems:
    - Intermittency
    - Reduced sensation
    - Waiting in the bathroom after voiding
    - Length of time in bathroom

- Planning life around toilet visits
- Social life
- Way feel about self

- The response options of the items ‘nocturia/nocturnal voids’, ‘daytime urinary frequency' and ‘length of time in bathroom’ were modified to mitigate floor effects.
- Three items were removed: ‘bowel associations’, ‘clustering of symptoms’ and ‘embarrassment’.
- Two items were added; ‘straining (to maintain)’ and ‘how often not getting enough sleep’. Further testing of the psychometric properties of these new items will occur in alongside planned clinical trials. Their responsiveness to change will provide further information on whether both items will be retained in the final instrument.
- The following items were flagged for potential removal, following the collection of sensitivity to change data:
  - Concentrates to start urination
  - Post-micturition dribble
  - Incontinence
  - Temporarily unable to pass urine
  - Physical activity
  - Fluid intake

6.6 Conclusion

The pilot study provided initial evidence to support the reliability and validity of both the ICIQ-UAB 1-week and the ICIQ-UAB 24hrs in patients with a diagnosis of DU with or without co-existing urological conditions. The evidence was used to make decisions regarding the removal or revision of items. Further psychometric testing (beyond the remit of this thesis) in a larger sample with the target clinical trial population will allow further evaluation of the instrument, including the assessment of sensitivity to change following an intervention and the derivation of a scientifically justified scoring system.

Chapter 7 Concept elicitation in the United States and Japan

7.1 Introduction

Direct patient involvement is essential in supporting aspects of content validity when developing items for a new PRO measure, to reflect the patients' perspective and when used to support labelling claims [99]. Considering the global intention for use of the ICIQ-UAB, a diverse sample as possible is required to ensure the eventual instrument represents the experiences of the demographic variations of the potential questionnaire respondents [112,216]. Special attention should be given to include the experiences of a patient population with diverse ethnic background during item development, in particular as the interviews in the UK were performed in a population of Caucasian patients. The following chapter describes the conduct of further concept elicitation interviews conducted in the United States (US) and Japan to support the content validity of the ICIQ-UAB instrument.

7.2 Methods

The objective was to elicit and document the experience of signs, symptoms and impact of UAB in patients diagnosed with DU, at sites in the US and Japan. The Japanese study was conducted at Nagoya University Hospital and Harasanshin Hospital. The study in the US was conducted at the Langone hospitals located in New York. Both studies were approved by the appropriate local ethics board in their respective countries.

In-depth qualitative interviews were conducted to explore patient's experience of UAB symptoms and the impact of those symptoms on their day-to-day lives. Patients with DU alone and those with other urological conditions were interviewed to capture the range of symptoms and impact which patients with UAB may experience. Patients recruited for the study met eligibility criteria based on the same urodynamic pressure flow study parameters as the interviews carried out in the UK.

Interviews were conducted by experienced qualitative researchers provided by contract research organisations, Pharmerit in the US, and IMS in Japan. Interviews were either face to face, or over the telephone. Written informed consent was obtained prior to the start of the interviews to audio record, transcribe verbatim and analyse the transcripts. The interviews

were guided by the interview schedule that was developed during the UK concept elicitation interviews. This was to enable the direct comparison between findings from the US, Japan and the Bristol study samples to help the documentation of conceptual adequacy across cultures [216]. The schedule was designed to be conducted in a way to encourage the patient to be as open as possible about their experiences in order to spontaneously elicit (without prompts) the symptoms and impact of their condition. Open-ended questions were followed by more targeted probes if required. The interviews took approximately 60 minutes.

7.2.1 Data analysis

Following data collection, the analysis of the transcripts was performed using the same approach and coding frame which was used in the UK concept elicitation study [22]. This entailed an initial inductive approach [134] to the data, which allowed the identification of 'new' symptoms or impacts if present, then transcript content was categorised or 'coded' by existing defined urological symptoms (e.g. 'hesitancy', 'urgency' and 'increased daytime frequency') [201,202]. In particular, if new concepts were mentioned spontaneously, or described more than once by subjects, these were flagged as potentially new concepts relevant to the respective population. All the US transcripts were coded by the author and independent qualitative researchers in the US and reconciled for differences following discussion meetings. For pragmatic reasons the Japanese interviews were conducted and analysed in the language of origin by independent Japanese qualitative researchers. The author was responsible for the reconciliation of any coding differences after translation in further correspondence with the Japanese researchers.

7.3 Results

7.3.1 Sample demographic and urodynamic characteristics

A total of 21 patients were interviewed, 10 from Japan and 11 in the US. The demographic and clinical characteristics of the sample from each country are shown in Table 32. A variety of educational backgrounds (ranging from high school to college or university educated), marital (single, married, divorced) and employment statuses (student, retired, employed) were represented in both population samples. In the US, all patients were Caucasian with the exception of one who was native Indian/Alaskan. A history of or current self-catheterisation, and the presence of a post void residual was common to many of the subjects in both samples. Participants all demonstrated DU, with or without urological co-existing conditions, as determined by pressure flow studies.

Table 32. Sample demographic and clinical characteristics.

Demographic or clinical characteristic	U.S. (n=11)	Japan (n=10)
Age		
Mean	60	66
SD	18.6	11.5
Range	24-86	39-78
Gender (n)		
Male	10	7
Female	1	3
Ethnicity (n)		
Caucasian/White	10	
Indian/Alaska native	1	
Japanese		10
Education level (n)		
High school	5	4
College or university	6	6
Marital status (n)		
Single	2	3
Married	7	7
Divorced	2	
Employment status (n)		
Student	1	
Employed (part-time or full-time)	6	3
Retired	4	2
Homemaker		2
Unemployed		3
Diagnostic group		
DU	7	6
DU + BOO-E	1	4
DU + SUI	1	
DU + DO	2	
PVR >30ml* *n* (%)	9 (82)	7 (70)
PVR (median and interquartile range) (ml)	230 (163-900)	162 (41-217)
BCI (median and interquartile range)	73 (38-92)	62 (49-75)
BOOI (median and interquartile range)**	0 (-1.2-11)	21 (16-27)
$P_{det}Q_{max}$ (cmH20) (median and interquartile range)	18.4 (16-24)	34 (19-37)
Q_{max} (ml/sec) (median and interquartile range)	14 (4-15)	5.5 (3.3-7)

*In the absence of any evidence base for the lower limit of a 'significant' PVR >30mls was chosen.

**Males only

Abbreviations: Detrusor pressure at maximum flow ($P_{det}Q_{max}$), maximum flow rate (Q_{max}), bladder contractility index (BCI) calculated by $BCI=P_{det}Q_{max}+5Q_{max}$, Bladder Outlet Obstruction Index (BOOI) calculated by $BOOI=P_{det}Q_{max}-2Q_{max}$, Post Void Residual (PVR).

7.3.2 Concept elicitation results

The following describes the main findings with relevant patient quotes derived from the original transcripts.

Storage symptoms

In the US sample, nocturia was the most commonly reported storage symptom; most patients (n=7/11) described having to get up at least once in the night to urinate. A high urinary frequency was often reported (n=4/11), however, the time between successive urinations was up to 12 hours in one patient. A reduced ability to tell when the bladder was full was reported (n=6/11). *US 201: "I don't have strong urges to empty my bladder".* Patients (n=5/11) described occurrences of a sudden desire to pass urine, which they were unable to ignore (urgency). Two participants described occasional stress or urgency urinary incontinence in small volumes: *"If I sneeze or if I cough...there's also urination going on at the same time."* There was one report of nocturnal enuresis.

In the Japanese sample, a high urinary frequency was reported by all of the patients (n=10/10). The number of micturitions was variable but was up to 13 times a day in one patient. Nocturia events of several times a night were also commonly reported (n=8/10). Urgency was described by many of the patients (n=7/10): *"I get this sudden desire to urinate even though I feel like there's no urine to pass"*. Some participants (n=6/10) described urinary incontinence in the context of sneezing, doing physical activity or when *"touching something cold".* A reduced sensation that the bladder was full was reported by a single participant: *"I feel abdominal pain when my bladder is full…but once I forget about it, I would not feel any sensation…"*

Voiding symptoms

In the US, the flow rate was usually described as *"weak"* (n=7/11) and could be very bothersome: *"At its worst, it was practically down to a trickle".* The symptom of straining was described as "pushing" or "squeezing" and was used when starting, maintaining or when finishing their urinations to try and empty their bladder as much as possible. Patients described (n=4/11) that it would *"take a little while to get going"* before the flow of urine would start (hesitancy). This could last for a few seconds to several minutes, dependent on the individual and the context. An intermittent stream (n=4/11) and urinations of small volume per void (n=4/11) were also reported.

Japanese patients (n=9/10) reported their urinary stream as having "weak force", especially in contrast to performance in the past. One patient emphasised the bothersome long urination time: *"The time it takes to complete urination is longer than it used to be".* Straining was also reported by most patients (n=8/10); this was used to initiate or finish urination, or to restart due an intermittent stream: "*I try to apply a lot of strength around my stomach".* A small volume per void (n=8/10) was commonly reported in the context of the other voiding symptoms. The symptom of hesitancy (n=8/10) was often described particularly in association with when they had been holding their urine in for a while: *"It doesn't come out. I would wait for 20, 30 minutes…I would sit and wait".* The splitting or spraying of the urinary stream was reported by three patients.

Post micturition symptoms

In the US, the sensation of incomplete emptying was commonly described (n=6/11). As a consequence, there was a need to return and pass urine a short time after (n=3/11), or to spend longer in the bathroom to ensure the bladder was as empty as possible: *"Whenever I go to the bathroom, I have to allow myself time... If I'm in a hurry, then I always regret it because I'll have to go again like 10 minutes later".* A post-micturition dribble was frequently reported in the sample (n=5/11).

Almost all the Japanese patients reported a sensation of incomplete emptying (n=9/10): "*I would always have that sensation that I* haven't completed released everything". A post-micturition dribble was reported by six of the patients: *"After I go to the toilet... a little bit of urine leaked out…".*

Other signs or symptoms

In the US, historical urinary tract infections were commonly reported (n=6/11) as well as a history of self-catheterisation (n=5/11). Incidents of acute retention requiring medical intervention were reported by several patients (5/11). On other occasions they were able to return to pass urine successfully a short time after (4/11). A small number of patients described a *"dull ache"* or a *"sharp, uncomfortable pressure"* in the bladder area (n=3/11). A minority of patients had bowel issues (n=3/11) which they associated with their urinary symptoms.

Japanese patients described the occurrence of bladder discomfort as a *"dull pain"*, or as a *"heavy"*, *"bloated"* sensation when the bladder was full or when they held their urine in

(n=8/10). A sensation of *"coldness"* in the bladder was also reported by one patient. Four patients currently or historically self-catheterised but urinary tract infections were not mentioned.

Impacts

In the US, the impact of their symptoms was variable between individuals. Some had adapted their lives around their condition and experienced relatively little impact (n=5/11). Others experienced a number of effects on day to day life including a reliance on knowing the location of toilets (n=7/11), tiredness from disturbed sleep (n=5/11) and disruption to social or daily activities including their work-life (n=3/11). The consequence was an impact on self-image or identity (n=6/11), and on relationships with family and friends (n=4/11). The emotional impact could be quite severe, resulting in depression in some cases: *"I was pretty depressed about it for a while, definitely felt like it affected my manhood in a way".* The practicalities of managing their symptoms (e.g. self-catheterisation) often led to feelings of anxiety or embarrassment in certain situations. Several mentioned the detrimental impact on their sex-lives (n=4/11), including a reduction in sex-drive: *"As a single person, losing the ability to have sex – that's a significant issue".* Some described having to carefully manage fluid intake to minimise disruption to sleep and their daily lives (n=5/11). Additional worry was also described due to the financial burden of their condition (n=3/11), as a result of increased medical expenses and complications securing medical insurance.

In Japan, although some of the patients reported little or no perceived impact on their lives, the planning of daily activities around the location of toilets such as when travelling, during recreational activities or when at work was also reported by most patients (n=8/10). Sleep disturbance and tiredness during the day due to nocturia was also a common impact (n=7/10). For many, there was a negative impact on how they felt about themselves (n=5/10): *"I don't know how women feel…but this condition is unattractive for me".* Feelings of anxiety, embarrassment or frustration were common due to their high urinary frequency, or if they were unable to access a toilet when required (n=5/10): *"I feel embarrassed about visiting the toilet multiple times".* Four of the patients described how their high urinary frequency affected the completion of physical activities such as hobbies, sport and household tasks. Maintaining personal hygiene was an issue for two patients: "*Because of the slow stream, my trousers get stained…I feel conscious about that."* An impact on sex-life and finances was not mentioned in this sample.

7.4 Discussion

To our knowledge, these studies represent the first qualitative exploration of the patient experience of UAB carried out in the United States or Japan. The findings from these studies increase the knowledge of the symptoms and impacts associated with UAB in a wider ethnic and cultural context.

The interviews in both countries elicited symptoms and impacts that support the previous findings from concept elicitation in the UK [22] and in the literature [21,32]. Storage symptoms of nocturia, high urinary frequency, urgency and incontinence and voiding symptoms of a slow and intermittent stream, hesitancy, straining and small volume per void were frequently reported by patients. Other symptoms reported were a sensation of incomplete emptying, a post-micturition dribble, a reduced sensation of bladder fullness, and bladder discomfort/pain. The signs of history of self-catheterisation and presence of a post void residual were also common to both samples. There were no symptoms reported that were unique to the US study, when compared to the Japanese and UK findings. However, US patients reported the financial impact of their condition, due to increased healthcare costs and health insurance premiums. The emotional impact was prominent particularly in the US sample (e.g. anxiety, effects on self-esteem and confidence) and several reported a negative impact on their sex-lives. In contrast, several of the Japanese patients reported the additional symptoms of bladder discomfort; a “bloated”, “heavy” sensation or “coldness”. Impacts that were specific to the Japanese population included frustration with their symptoms, the ability to maintain good personal hygiene, and the impact on the appearance of their clothing.

There were a number of findings that have implications for the cultural adaptation of the UAB PRO instrument. The increased prominence of the reporting of impact on sex-lives in the US sample could reflect cultural openness of talking about sex-related issues. However, it is recognised that side-effects of urological surgery or medication which can affect sex-drive and/or erectile dysfunction in men [217] complicates measurement of this concept. Additional items or modification to the existing items in the ICIQ-UAB which measure the emotional impact (e.g. frustration, anxiety, effects on self-esteem and confidence) could be considered. The financial impact of UAB reported in the US is unlikely to be included as an additional item as this concept is applicable to any disease requiring healthcare, dependent on income, and would not respond to a healthcare intervention. Although coded as ‘lower urinary tract pain’ in the UK interviews, the character of any bladder discomfort or pain experienced has perhaps been clarified further by this study. Pain or discomfort, perhaps as a result of a post void

residual or full bladder appears to be a concept that is reported frequently, however, the location or underlying aetiology of this pain remains uncertain. A long urination time and impacts of hygiene and on clothing appearance in the Japanese sample were reported by a minority of patients. When making decisions surrounding the modification, inclusion or removal of items, consideration must be given to the frequency, bother, and the spontaneity by which a concept was reported, as well as their clinical utility.

Limitations

Qualitative findings, although transferable, should be interpreted within the context in which they were obtained. This study is not intended to provide accurate estimates of symptom and impact frequency. The intention was for the experiences to be representative, not the sample population [116].

Despite significant efforts made, no African American patients or Hispanics were recruited in the US. However, there was a good spread of demographic characteristics and ages recruited which may be typical of men and women who present and are referred for urodynamics in these countries.

Underlying DU is known to have a number of possible aetiologies which may be myogenic, neurogenic or idiopathic in origin [24,46,57]. Further research is required to establish whether there are symptomatic differences in presentation due to classifications of DU by aetiology. It is also acknowledged that the underlying aetiology of frequently reported symptoms such as nocturia and bladder pain are complex [47,218], which has implications for the potential responsiveness of associated items within the UAB PRO measure to an intervention.

7.5 Instrument modification following the interviews in the US and Japan

Following evidence from the concept elicitation interviews in US and Japan and discussion with clinical study team, some changes were made (to ICIQ-UAB v13 resulting in version 14), which were tested by four confirmatory cognitive interviews. The details and rationale for these changes are given in Appendix 8. Final decisions implemented in the developmental ICIQ-UAB (version 15) are summarised as follows:

- Minor changes to the initial instructions and daytime urinary frequency item response options.
- A dichotomous yes/no item was added to capture historical self-catheterisation.

- The item relating to incontinence item was split into two items, to capture and differentiate stress versus urgency incontinence.
- An item was adapted from the original cognitive interviews in the UK and re-instated to capture bladder pain.

The final 'developmental version' ICIQ-UAB (v15) is given in Appendix 9.

7.6 Conclusions

The findings in the US and Japan largely corroborate the findings obtained in the UK and explores the concepts among alternative populations. This supports the evidence for the ICIQ-UAB as a tool that can be used globally. These studies also broaden our knowledge of the symptoms and impacts associated with UAB in wider ethnic and cultural backgrounds, and provided valuable input to the further refinement of the developmental questionnaire. The results informed decisions surrounding item modification for the development of the PRO measure.

Chapter 8 Summary and conclusions

8.1 Introduction

The developmental version of the ICIQ-UAB described in this thesis is the culmination of a comprehensive set of sub-studies to provide data for its initial validation. The validity, reliability and wider cultural applicability of the ICIQ-UAB among men and women with a confirmed diagnosis of DU are supported by the findings from interviews in the UK, US, and Japan, and European pilot psychometric testing. The ICIQ-UAB is the first PRO measure for the assessment of the symptoms and impact of UAB which has been developed using rigorous methodology, in-line with the FDA Guidance for Industry [99]. Following further planned evaluation of the instrument's responsiveness to change and the derivation of a scoring system, the instrument is envisaged as an important tool for the monitoring of future treatment strategies for patients with UAB. Extracts from the following summary of the development activities to date were submitted as a manuscript to the Journal of Neurourology and Urodynamics in May 2018.

8.2 Summary of development activities

The review of the contemporary literature identified more than twenty PRO instruments with published evidence of psychometric properties, that evaluated storage, voiding and post-micturition LUTS and/or associated health related quality of life. Although existing PRO measures capture some of the symptoms reported by patients with UAB, these instruments were developed for populations of individuals with broad LUTS, which defines their context of use. The inclusion of items, including the wording and content of the PRO instruments, should be based on the patient-centred input of the specific population of interest [99,112]. The lack of an existing, psychometrically robust PRO instrument for the assessment of UAB developed to current regulatory standards justified the development of the new instrument.

The initial qualitative phase of development with patients from the target population, alongside the consultation with an expert clinical panel resulted in a comprehensive instrument which included the items of relevance, was interpreted as intended, and easy to complete. The concept elicitation interviews elucidated the patient experience of the condition to identify pertinent items, including the descriptive wording of their symptoms and

their impact, day to day variability, severity and bother. The results of this qualitative study were published in European Urology [22]. More than 20 LUTS and signs with associated impact on quality of life were described by the patients with DU. Descriptions of the patient experience of UAB were consistent with the known symptoms associated with DU in the literature [15,19,32] and symptomatic definition of UAB [27]. A draft instrument was generated following the UK concept elicitation interviews. Rounds of cognitive interviews were then scheduled and iterative improvements made until all items were considered to be fully understood and interpreted as intended by the respondents. Two revised versions of the ICIQ-UAB, one with a recall period of 1-week and the other with a recall period of 24hrs were considered comprehensive and ready for the next phase of development involving the psychometric testing of its properties.

The pilot testing provided encouraging evidence of the validity and reliability of the draft instrument. Test-retest reliability in stable patients was good (ICC ≥ 0.85 for both versions), providing evidence of the reliability of scores over the test period of 10 days, and that any variability was not due to systematic differences among respondents [104]. The score derived from the tested domains in both versions demonstrated reliability ($\alpha \geq 0.85$) which is over the accepted threshold of ≥0.7 [154] for internal consistency. Many of the symptom and impact items in both versions exhibited floor effects, which can affect the instrument's sensitivity to change [99]. However, these effects could also be capturing the true severity or impact of the symptoms in the studied population. The known group validity of the items was supported for both recall period versions, so the instrument score was shown to be sensitive to severity of condition. Construct validity was demonstrated by the expected convergent and divergent correlations with other PRO measures of known validity. There was no strong evidence from the pilot study to suggest an overall advantage for either recall period version. Going forward, a shorter recall period is recommended by the FDA guidance for industry to reduce recall bias [99], but this should be carefully considered to match the condition, the PRO domain being measured and views of patients [219]. For example, we know from the patient interviews there was a preference for a longer recall period to reflect the variation of some symptoms and in particular the required length of time to allow a measureable impact.

The additional concept elicitation interviews carried out in the US and Japan represented the first qualitative exploration of the patient experience of UAB with patients confirmed to have DU in these countries. The interviews elicited symptoms and impacts that support previous qualitative findings in the UK [22] and the overall content validity of the ICIQ-UAB instrument.

The evidence from the pilot study and the interviews in the US and Japan were used to inform modifications to the instrument. These changes were tested for their acceptability by two sets of additional cognitive interviews, resulting in the final 33-item 'developmental version'.

A flow diagram illustrating the developmental progression of the ICIQ-UAB is given in Figure 11.

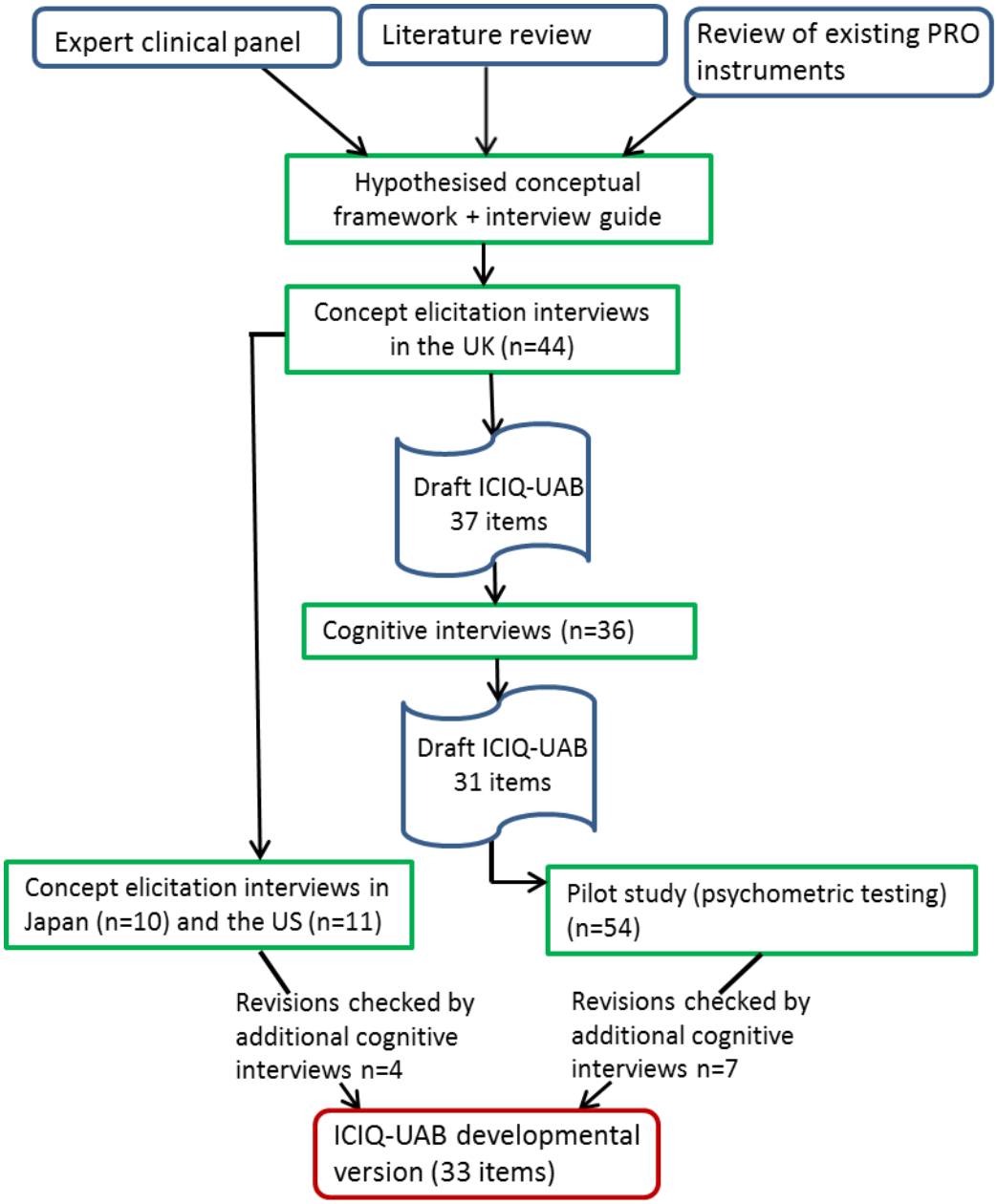


Figure 11. The development of the ICIQ-UAB.

8.3 Further development activities

8.3.1 Sensitivity to change

The properties of validity and reliability have been supported by the sub-studies detailed. However, responsiveness to change to an intervention has not yet been assessed. As discussed in chapter 1, the treatment options for patients with DU are limited, which makes the assessment of sensitivity to change problematic. Intermittent self-catheterisation is currently the only intervention available to patients, other than the currently unsatisfactory and often ineffective surgical, pharmaceutical and electrostimulation options.

ISC is known to be successful in alleviating symptoms for neuropathic bladder patients and for bladder drainage [211], but any apparent treatment benefit for patients with UAB are not well established [25]. Nevertheless, incomplete bladder emptying is considered an indicative symptom of UAB [26] and a high proportion of UAB patients have a post void residual [22,32] so it is reasonable to expect some treatment benefit from learning ISC for UAB patients. Indeed, there was some evidence from the UK concept elicitation interviews that some relief of storage symptoms may occur from practising ISC. Thus, it is hypothetically possible the ICIQ-UAB may be sensitive to change following an ISC intervention.

Two test patients with suspected DU (not urodynamically confirmed) were given the ICIQ-UAB in clinic before learning ISC (1st administration) and then again four weeks later at their follow-up appointment (2nd administration). The Patient Global Impression of Improvement (PGI-I) was also given at the 2nd administration. Their total scores for symptom, impact and bother domains are shown in Table 33. The scores for selected storage and post-micturition symptom items are given in Table 34.

Patient 1 reported 'no change' in their symptoms on the PGI-I at the 2nd administration, and patient 2 reported that their symptoms were 'a little worse'. In both patients there was a small increase in symptom, impact and bother scores from the 1st administration to the 2nd administration after learning ISC. At the item level (Table 34), the reported scores in the 2nd administration for the items relating to urgency, nocturnal voids and daytime urinary frequency were higher, but scores for incontinence, need to immediately re-void and sensation of incomplete emptying were slightly lower. The other symptom items had a difference of no more than +/-1.

Table 33. ICIQ-UAB scores of two test patients before and after learning ISC.

	Symptom score		Impact score		Bother score (symptoms)		Bother score (impacts)	
Administration	1st	2nd	1st	2nd	1st	2nd	1st	2nd
Patient 1	33	36	4	4	69	88	16	17
Patient 2	55	59	16	18	152	169	55	63

Table 34. ICIQ-UAB scores of selected storage and post-micturition items.

Item	**Administration item total score (n=2)**		**Score difference**
	1st	**2nd**	
Urgency (Q16)	1	4	3
Nocturnal voids (Q17)	2	4	2
Daytime urinary frequency (Q18)	1	2	1
Hesitancy (4)	6	8	2
Incontinence (Q8)	1	0	-1
Need to immediately re-void (Q9)	4	3	-1
Sensation of incomplete emptying (Q10)	7	6	-1

Although no conclusions on sensitivity to change can be made due to the test sample size, it is evident further research is required regarding the measurement capabilities of the ICIQ-UAB and on the efficacy of learning ISC for patients with suspected underactive bladder. It is also important to recognise that any symptom alleviation as a result of ISC must outweigh the known inconvenience of self-catheterisation on day-to-day life [22]. Although the ICIQ-UAB was tested during the qualitative phases of its development with patients who regularly self-catheterise, it is acknowledged that items that ask about voiding symptoms are not applicable if a patient self-catheterises for every void. This may reduce the sensitivity of the ICIQ-UAB to change as a result of an ISC intervention.

The ICIQ-UAB was designed and developed according to FDA guidelines [99] with the intention of its eventual use as a primary clinical outcome measure in clinical trials. It is anticipated that responsiveness to an intervention will be assessed in a phase 2 trial of a new drug. The ICIQ-UAB will be administered before and after the intervention and the scores correlated with other measures (pad tests, bladder diaries, other validated PRO measures). The ICIQ-UAB is currently designed to comprehensively cover all concepts including 30 items, but it is anticipated that there is scope for removal of items following further data collection on their properties. The item-level analysis of validity, reliability and responsiveness alongside future

clinical trials will inform decision making for removing poorly performing items. The resulting tool will then potentially be used in phase III trials as a primary outcome measure.

It is anticipated that alongside responsiveness to change, data from future clinical trials will be analysed using factor analysis to explore the PRO's structure, possible scoring system, and to assess suitability for the application of item response theory (IRT) techniques. IRT modelling provides additional information on the item properties and is particularly useful when making decisions on to reduce the number of items in the questionnaire [166]. The use of these techniques requires a sample size of several hundred so are often used in the later stages of psychometric testing.

8.3.2 Electronic equivalence and application

The ICIQ-UAB has been developed and validated for pencil-and-paper administration. However, a questionnaire which can be shown to be reliable in different administrative formats has further potential for utility in different contexts [220]. There are a number of potential advantages to electronic data capture when collecting PRO data. These include the reduction of administrative data entry workload, greater accuracy and completeness of data [221–223]. Patient acceptance of using tablet computers to complete questionnaires is generally high which can also increase compliance [224–226]. In addition, a questionnaire which has been validated for equivalence over the telephone can provide additional flexibility for delivery [227–229].

The conversion to electronic patient reported outcomes (ePROs) requires the demonstration of equivalence to the paper-and-pencil version [230,231]. Scores cannot be assumed to be equivalent [232], and must not differ simply due to the method of data collection that is used. Uren et al. (2017) aimed to evaluate whether scores obtained from patient-completed entry of four different ICIQ PRO questionnaires on a touch screen device (iPad) were equivalent with corresponding data collected using conventional pencil-and-paper methods, or when administered over the telephone [233]. A total of 491 men and women, attending the Bristol Urological Institute complaining of LUTS, were randomised to one of three groups which determined the order in which they completed three administrations of the same questionnaire: paper, iPad and telephone. Four ICIQ questionnaires were evaluated: ICIQ-MLUTS, ICIQ-LUTSqol, ICIQ-OABqol and ICIQ-UI SF. The results showed that iPad and paper-and-pencil administrations of the ICIQ modules tested produced scores that were equivalent. This was demonstrated by very high overall correlations of the scores obtained by the iPad

administered version with those obtained by repeated paper administrations. Paper and phone administrations were less well correlated, although still high, but this may reflect the requirement for a proxy required when using the telephone. The ICIQ-UAB formatting, layout and response options were based on the existing ICIQ modules so it is reasonable to generalise that scores obtained by patients in an electronic version would have the same level of equivalence [230,234]. Indeed, current evidence recommends that full psychometric validation may be unnecessary for minor modifications to questionnaire format [230,234]. This study provides the validation required for the development and use of an application or 'app' to be made available for patients to be able to complete electronic versions of the ICIQ-UAB on mobile touch screen devices. Alongside the other modules of the ICIQ, this significantly increases the versatility of the ICIQ-UAB for use in both clinical and research settings.

In addition it may be possible to use computer adaptive testing (CAT) to select the most appropriate items during administration [172]. "CAT is a computer algorithm that selects successive items based on an individual's responses to previously administered items" [170]. For example, if the first item asks 'how often do you have to get up and urinate each night, on average?' and the respondent' answer was 5 times, this may suggest that that underlying severity of UAB may be high. The CAT algorithm could then select another question that allows greater precision for those who lie in the severe range of population. All items are calibrated on to a common scale so scores can then be compared even if different items were used, and items may be added or removed without compromising the validity of the overall instrument. Thus, the precision is optimised for the given questionnaire length and irrelevant items are excluded [235]. The precision of the test can also be set to match the needs of the instrument, for example high precision may be required for diagnostic purposes.

These are exciting new directions that could considerably expand the versatility of the ICIQ-UAB upon completion of the final version. Early awareness of the intended techniques to be considered with the questionnaire is useful to prospectively plan future studies and evaluation efforts.

8.4 Strengths and limitations

It could be argued that it is premature to develop a PRO questionnaire for underactive bladder when the symptomatic definition is not yet fully established. Indeed, a symptomatic definition is not easy to define, as there are no distinguishing symptoms which encompass the underactive bladder syndrome. The UAB presenting symptoms also occur as a result of other

urological conditions with entirely different aetiologies, such as bladder obstruction. However, the most recent ICS definition has been accepted for publication in Neurology and Urodynamics in May 2018, and this area continues to be the subject of considerable attention. Recent research also continues to reveal symptomatic differences in presenting symptoms between those with DU and those with DU and co-existing other urological conditions [21,236]. The concept of an 'underactive bladder' is one that patients can relate to as a perceived weakness of the bladder muscle, manifesting as a reduced flow rate and incomplete emptying. For clinicians, it is a useful concept to describe the condition, and one which enables the patient understanding of the different urological conditions with distinct aetiologies but similar presentation. The choice of further treatment is then easier to explain to the patient, rather than having to resort to technical terminology, such as a diagnosis of DU. Thus, the ICIQ-UAB has been developed as a useful tool to assess those with suspected or confirmed DU, for the further assessment of their symptoms in response to conservative or other suggested interventions or monitor symptom status over time, in addition to its widespread research potential.

What information can a PRO measure for underactive bladder provide which other objective tests, such as bladder diaries, pad tests or uroflowmetry cannot? Parameters such as the frequency of urination, flow rate, and nocturnal urinary episodes can be collected by these tests as an adjunct to the questionnaire data. However, PRO measure data captures subjective information which only the patient can provide. If well validated, a PRO questionnaire provides a way of capturing subjective information in a structured, comprehensive and interpretable way. The clinician can then see which symptoms are particularly bothersome to the patient, which is useful not only for diagnostic purposes but for the tailoring of possible treatment strategies.

Overall strengths of the study include the recruitment from multiple sites in Europe, the US, and Japan; the strict adherence to *a priori* urodynamic inclusion criteria, and the first conduct of robust psychometric PRO development methodology in the target population with UAB/DU.

A total of 143 individual patients were recruited into all sub-studies. Of these, a total of 65 qualitative concept elicitation interviews were conducted with patients recruited at multiple sites in the UK, US and Japan. It may be considered very likely that the major symptoms, and behaviours presented by the target population with underactive bladder have been elicited by these interviews. Indeed, data saturation tables for the UK and US interviews confirmed that saturation of concept had been achieved. Attempts were made for the sample to be as

representative as possible of the presenting population with UAB by the purposive sampling of the views of possibly underrepresented groups. Ethnic minorities were particularly challenging to recruit, however, the input of culturally diverse populations from the US and Japan provided further insight into the final choice of items and their wording. The epidemiology of the background population of underactive bladder is poorly understood, so the sample is more likely representative of the referred populations in participating sites rather than the background population. Indeed, the lower number of females may be a consequence of the referred nature of the sample as there is a greater necessity for diagnostic pressure flow studies performed prior to prostatectomy to differentiate BOO from DU in men [203]. The overall sample characteristics such as the mean age, high PVR and the reported symptoms were comparable to recent publications [21] and are consistent with the recent ICS symptomatic definition of UAB.

The size of the pilot test sample was larger than the recommended minimum size of 40 to obtain reliable estimates of item performance and reliability [206]. However, decisions made on the basis of this evidence were purposely conservative; The ICIQ-UAB is still relatively lengthy at 33 items, as it is advantageous to retain all concepts of possible interest to the population at this stage until further evidence of item performance is collected. The item-level analysis, validity and reliability will be further evaluated alongside responsiveness to change, using data from future clinical trials.

The ICIQ-UAB is designed as an outcome measure and not as a diagnostic tool for DU/UAB, due to the overlap of reported symptoms with co-existing urological conditions. Underlying DU is known to have a number of possible aetiologies and may be myogenic or neurogenic in origin [24]. Further research is required to establish whether there are symptomatic differences in presentation due to classifications of DU by aetiology. It is also acknowledged that the underlying aetiology of frequently reported symptoms such as nocturia and bladder pain are complex [47,218], which could affect the responsiveness to an intervention of the associated items. It is anticipated that when more data is acquired through a larger sample in future clinical trials, the symptom severity and prevalence will be further elucidated relative to other urological conditions.

8.5 Summary of future research and direction for the ICIQ-UAB

As concluded in chapter 1, further research is required in understanding the epidemiology, aetiology and pathophysiology of underactive bladder and detrusor underactivity. This will

influence the clinical applicability of the ICIQ-UAB and uptake into clinical practice. Thus, recommendations for research and the future direction of the ICIQ-UAB are intrinsically linked. Consensus of the most recent symptomatic definition has also only just been achieved. Further professional consensus of urodynamic parameters that characterise DU for men and women are also required to enable multidisciplinary and wider global recognition of the condition.

Recommendations for the future direction for the ICIQ-UAB:

- The exploration of an underlying domain structure by factor analysis to explore the unidimensionality of the underlying constructs and to develop appropriate scoring systems.
- The sensitivity to change is to be evaluated in a larger population, as part of planned clinical trials.
- The development and use of an electronic version of this and other ICIQ modules.
- The formation of an ICIQ item bank and the possible use of computer adaptive testing to increase the versatility of the ICIQ modules for global users.
- The improvement of scoring algorithms and PRO accuracy using IRT modelling where appropriate, with larger datasets.

8.6 Conclusion

The ICIQ-UAB has been developed and tested with symptomatic patients diagnosed with detrusor underactivity, with or without co-existing urological conditions. The result is a novel developmental instrument, applicable to the heterogeneous target population which is encompassed by the symptom complex of underactive bladder. The content validity of the instrument is evidenced by rigorous qualitative methodology and patient input, according to the latest regulatory guidelines. The initial quantitative evaluation of validity and reliability were also supportive of these psychometric properties. The development of the ICIQ-UAB was not without its challenges. There is only recent consensus at the level of its symptomatic definition, and the understanding of its aetiology, epidemiology and pathophysiology still requires much research. Treatment options are still very limited, with available options often ineffective or still awaiting clinical trials to evidence their efficacy. However, as the field develops there is the promise of more effective treatment options in the future. The ICIQ-UAB is envisaged as an important tool for the measurement of outcomes and treatment benefit for these new interventions, as it is continues its development as an outcome measure and eventual integration into clinical practice.

References

1. Bates C p., Whiteside CG, Turner-Warwick R. Synchronous Cine/Pressure/Flow/Cysto-Urethrography with Special Reference to Stress and Urge Incontinence. *Br J Urol*. 1970;42(6):714-723.

2. Uren AD, Drake MJ. Definition and symptoms of underactive bladder. *Investig Clin Urol*. 2017;58(Suppl 2):S61-S67.

3. Abrams P. New words for old: lower urinary tract symptoms for “prostatism”. *BMJ*. 1994;308(6934):929-930.

4. Abrams P, Cardozo L, Fall M, et al. The standardisation of terminology of lower urinary tract function: report from the Standardisation Sub-committee of the International Continence Society. *Neurourol Urodyn*. 2002;21(2):167-178.

5. Irwin DE, Milsom I, Hunskaar S, et al. Population-Based Survey of Urinary Incontinence, Overactive Bladder, and Other Lower Urinary Tract Symptoms in Five Countries: Results of the EPIC Study. *Eur Urol*. 2006;50(6):1306-1315.

6. Boyle P, Robertson C, Mazzetta C, et al. The prevalence of lower urinary tract symptoms in men and women in four centres. The UrEpik study. *BJU Int*. 2003;92(4):409-414.

7. Patel ND, Parsons JK. Epidemiology and etiology of benign prostatic hyperplasia and bladder outlet obstruction. *Indian J Urol*. 2014;30(2):170-176.

8. Coyne KS, Wein AJ, Tubaro A, et al. The burden of lower urinary tract symptoms: evaluating the effect of LUTS on health-related quality of life, anxiety and depression: EpiLUTS. *BJU Int*. 2009;103 Suppl 3:4-11.

9. Coyne KS, Sexton CC, Kopp Z, et al. Assessing patients’ descriptions of lower urinary tract symptoms (LUTS) and perspectives on treatment outcomes: results of qualitative research. *Int J Clin Pract*. 2010;64(9):1260-1278.

10. Abrams P. *Urodynamics*. Third Edition. London: Springer; 2006.

11. Wein AJ, Rovner ES. Definition and epidemiology of overactive bladder. *Urology*. 2002;60(5 Suppl 1):7-12; discussion 12.

12. Abrams P, Cardozo L, Fall M, et al. The standardisation of terminology in lower urinary tract function: report from the standardisation sub-committee of the International Continence Society. *Urology*. 2003;61(1):37-49.

13. Rovner ES, Goudelocke CM. Urodynamics in the evaluation of overactive bladder. *Curr Urol Rep*. 2010;11(5):343-347.

14. Chapple CR, Osman NI, Birder L, et al. The Underactive Bladder: A New Clinical Concept? *Eur Urol*. 2015;68(3):351-3.

15. Osman NI, Chapple CR, Abrams P, et al. Detrusor Underactivity and the Underactive Bladder: A New Clinical Entity? A Review of Current Terminology, Definitions, Epidemiology, Aetiology, and Diagnosis. *Eur Urol*. 2014;65(2):389-398.

16. Taylor JA, Kuchel GA. Detrusor underactivity: Clinical features and pathogenesis of an underdiagnosed geriatric condition. *J Am Geriatr Soc*. 2006;54(12):1920-1932.

17. Griffiths D. Detrusor contractility--order out of chaos. *Scand J Urol Nephrol Suppl*. 2004;(215):93-100.

18. Smith PP. Aging and the underactive detrusor: a failure of activity or activation? *Neurourol Urodyn*. 2010;29(3):408-412.

19. van Koeveringe GA, Vahabi B, Andersson KE, Kirschner-Herrmans R, Oelke M. Detrusor underactivity: a plea for new approaches to a common bladder dysfunction. *Neurourol Urodyn*. 2011;30(5):723-728.

20. van Koeveringe GA, Rademakers KLJ, Birder LA, et al. Detrusor underactivity: Pathophysiological considerations, models and proposals for future research. ICI-RS 2013. *Neurourol Urodyn*. 2014;33(5):591-596.

21. Gammie A, Kaper M, Dorrepaal C, Kos T, Abrams P. Signs and Symptoms of Detrusor Underactivity: An Analysis of Clinical Presentation and Urodynamic Tests From a Large Group of Patients Undergoing Pressure Flow Studies. *Eur Urol*. 2016;69(2):361-369.

22. Uren AD, Cotterill N, Harding C, et al. Qualitative Exploration of the Patient Experience of Underactive Bladder. *Eur Urol*. 2017;72(3):402-407.

23. Fode M, Sønksen J. Towards a Greater Understanding of Underactive Bladder. *Eur Urol*. 2017;72(3):408-409.

24. Dewulf K, Abraham N, Lamb LE, et al. Addressing challenges in underactive bladder: recommendations and insights from the Congress on Underactive Bladder (CURE-UAB). *Int Urol Nephrol*. 2017;49(5):777-785.

25. Miyazato M, Yoshimura N, Chancellor MB. The Other Bladder Syndrome: Underactive Bladder. *Rev Urol*. 2013;15(1):11-22.

26. Chapple CR, Osman NI. Crystallizing the Definition of Underactive Bladder Syndrome, a Common but Under-recognized Clinical Entity. *Low Urin Tract Symptoms*. 2015;7(2):71-76.

27. Wein A, Chapple C. Introduction and Terminology. In: Chapple C, Wein A, Osman N, eds. *Underactive Bladder*. 1st Edition. Switzerland: Springer; 2017:ix-xiii.

28. Osman N, Hillary C, Chapple C. Epidemiology of Underactive Bladder. In: Chapple C, Wein A, Osman N, eds. *Underactive Bladder*. 1st Edition. Switzerland: Springer; 2017:25-31.

29. Valente S, DuBeau C, Chancellor D, et al. Epidemiology and demographics of the underactive bladder: a cross-sectional survey. *Int Urol Nephrol*. 2014;46 Suppl 1:S7-10.

30. Abarbanel J, Marcus E-L. Impaired detrusor contractility in community-dwelling elderly presenting with lower urinary tract symptoms. *Urology*. 2007;69(3):436-440.

31. Jeong SJ, Kim HJ, Lee YJ, et al. Prevalence and Clinical Features of Detrusor Underactivity among Elderly with Lower Urinary Tract Symptoms: A Comparison between Men and Women. *Korean J Urol*. 2012;53(5):342-348.

32. Hoag N, Gani J. Underactive Bladder: Clinical Features, Urodynamic Parameters, and Treatment. *Int Neurourol J*. 2015;19(3):185-189.

33. Fusco F, Groutz A, Blaivas JG, Chaikin DC, Weiss JP. Videourodynamic studies in men with lower urinary tract symptoms: a comparison of community based versus referral urological practices. *J Urol*. 2001;166(3):910-913.

34. Groutz A, Gordon D, Lessing JB, Wolman I, Jaffa A, David MP. Prevalence and characteristics of voiding difficulties in women: are subjective symptoms substantiated by objective urodynamic data? *Urology*. 1999;54(2):268-272.

35. Valentini FA, Robain G, Marti BG. Urodynamics in women from menopause to oldest age: what motive? What diagnosis? *Int Braz J Urol Off J Braz Soc Urol*. 2011;37(1):100-107.

36. Wang CC, Yang SSD, Chen YT, Hsieh JH. Videourodynamics identifies the causes of young men with lower urinary tract symptoms and low uroflow. *Eur Urol*. 2003;43(4):386-390.

37. Nitti VW, Lefkowitz G, Ficazzola M, Dixon CM. Lower urinary tract symptoms in young men: videourodynamic findings and correlation with noninvasive measures. *J Urol*. 2002;168(1):135-138.

38. Cohn JA, Brown ET, Kaufman MR, Dmochowski RR, Reynolds WS. Underactive bladder in women: is there any evidence? *Curr Opin Urol*. 2016;26(4):309-314.

39. Kim M, Jeong CW, Oh S-J. Effect of Preoperative Urodynamic Detrusor Underactivity on Transurethral Surgery for Benign Prostatic Hyperplasia: A Systematic Review and Meta-Analysis. *J Urol*. 2018;199(1):237-244.

40. Pfisterer MH-D, Griffiths DJ, Schaefer W, Resnick NM. The effect of age on lower urinary tract function: a study in women. *J Am Geriatr Soc*. 2006;54(3):405-412.

41. Ameda K, Sullivan MP, Bae RJ, Yalla SV. Urodynamic characterization of nonobstructive voiding dysfunction in symptomatic elderly men. *J Urol*. 1999;162(1):142-146.

42. Malone-Lee J, Wahedna I. Characterisation of detrusor contractile function in relation to old age. *Br J Urol*. 1993;72(6):873-880.

43. Karram MM, Partoll L, Bilotta V, Angel O. Factors affecting detrusor contraction strength during voiding in women. *Obstet Gynecol*. 1997;90(5):723-726.

44. Madersbacher S, Pycha A, Schatzl G, Mian C, Klingler CH, Marberger M. The aging lower urinary tract: a comparative urodynamic study of men and women. *Urology*. 1998;51(2):206-212.

45. Fry CH, Bayliss M, Young JS, Hussain M. Influence of age and bladder dysfunction on the contractile properties of isolated human detrusor smooth muscle. *BJU Int*. 2011;108(2b):E91-E96.

46. Drake MJ, Williams J, Bijos DA. Voiding dysfunction due to detrusor underactivity: an overview. *Nat Rev Urol*. 2014;11(8):454-464.

47. Drake MJ. Should nocturia not be called a lower urinary tract symptom? *Eur Urol*. 2015;67(2):289-290.

48. Gulur DM, Mevcha AM, Drake MJ. Nocturia as a manifestation of systemic disease. *BJU Int*. 2011;107(5):702-713.

49. Gannon K, Glover L, O'Neill M, Emberton M. Men and Chronic Illness: A Qualitative Study of LUTS. *J Health Psychol*. 2004;9(3):411-420.

50. Glover L, Gannon K, McLoughlin J, Emberton M. Men's experiences of having lower urinary tract symptoms: factors relating to bother. *BJU Int*. 2004;94(4):563-567.

51. Wareing M. Lower urinary tract symptoms: a hermeneutic phenomenological study into men's lived experience. *J Clin Nurs*. 2005;14(2):239-246.

52. Birder L, de Groat W, Mills I, Morrison J, Thor K, Drake M. Neural control of the lower urinary tract: peripheral and spinal mechanisms. *Neurourol Urodyn*. 2010;29(1):128-139.

53. Fowler CJ, Griffiths D, de Groat WC. The neural control of micturition. *Nat Rev Neurosci*. 2008;9(6):453-466.

54. de Groat WC. Integrative control of the lower urinary tract: preclinical perspective. *Br J Pharmacol*. 2006;147 Suppl 2:S25-40.

55. Aizawa N, Igawa Y. Pathophysiology of the underactive bladder. *Investig Clin Urol*. 2017;58(Suppl 2):S82-S89.

56. Kadow BT, Tyagi P, Chermansky CJ. Neurogenic Causes of Detrusor Underactivity. *Curr Bladder Dysfunct Rep*. 2015;10(4):325-331.

57. Osman NI, Chapple CR. Contemporary concepts in the aetiopathogenesis of detrusor underactivity. *Nat Rev Urol*. 2014;11(11):639-648.

58. Pizzi A, Falsini C, Martini M, Rossetti MA, Verdesca S, Tosto A. Urinary incontinence after ischemic stroke: clinical and urodynamic studies. *Neurourol Urodyn*. 2014;33(4):420-425.

59. Litwiller SE, Frohman EM, Zimmern PE. Multiple sclerosis and the urologist. *J Urol*. 1999;161(3):743-757.

60. Amarenco G, Raibaut P, Hubeaux K, Jousse M, Sheikh Ismaël S, Lapeyre E. [Autonomic nervous system alteration in multiple sclerosis patients with urinary symptoms. Clinical, urodynamic and cardiovascular study]. *Progres En Urol J Assoc Francaise Urol Soc Francaise Urol*. 2013;23(17):1505-1510.

61. Araki I, Matsui M, Ozawa K, Takeda M, Kuno S. Relationship of bladder dysfunction to lesion site in multiple sclerosis. *J Urol*. 2003;169(4):1384-1387.

62. Elbadawi A. Pathology and pathophysiology of detrusor in incontinence. *Urol Clin North Am*. 1995;22(3):499-512.

63. Nordling J. The aging bladder--a significant but underestimated role in the development of lower urinary tract symptoms. *Exp Gerontol*. 2002;37(8-9):991-999.

64. Andersson KE. Pathophysiology and Associations of Underactive Bladder. In: *Underactive Bladder*. Springer, Cham; 2017:1-12.

65. Al-Hayek S, Thomas A, Abrams P. Natural history of detrusor contractility--minimum ten-year urodynamic follow-up in men with bladder outlet obstruction and those with detrusor. *Scand J Urol Nephrol Suppl*. 2004;(215):101-108.

66. Thomas AW, Cannon A, Bartlett E, Ellis-Jones J, Abrams P. The natural history of lower urinary tract dysfunction in men: minimum 10-year urodynamic follow-up of untreated detrusor underactivity. *BJU Int*. 2005;96(9):1295-1300.

67. Ponholzer A, Temml C, Wehrberger C, Marszalek M, Madersbacher S. The association between vascular risk factors and lower urinary tract symptoms in both sexes. *Eur Urol*. 2006;50(3):581-586.

68. Tarcan T, Azadzoi KM, Siroky MB, Goldstein I, Krane RJ. Age-related erectile and voiding dysfunction: the role of arterial insufficiency. *Br J Urol*. 1998;82 Suppl 1:26-33.

69. Nomiya M, Yamaguchi O, Akaihata H, et al. Progressive vascular damage may lead to bladder underactivity in rats. *J Urol*. 2014;191(5):1462-1469.

70. Azadzoi KM, Tarcan T, Kozlowski R, Krane RJ, Siroky MB. Overactivity and structural changes in the chronically ischemic bladder. *J Urol*. 1999;162(5):1768-1778.

71. Chancellor MB. The overactive bladder progression to underactive bladder hypothesis. *Int Urol Nephrol*. 2014;46 Suppl 1:S23-27.

72. Fedele D. Therapy Insight: sexual and bladder dysfunction associated with diabetes mellitus. *Nat Clin Pract Urol*. 2005;2(6):282-290; quiz 309.

73. Verrotti A, Prezioso G, Scattoni R, Chiarelli F. Autonomic Neuropathy in Diabetes Mellitus. *Front Endocrinol*. 2014;5.

74. Andersson K-E. Bladder underactivity. *Eur Urol*. 2014;65(2):399-401.

75. Klingler HC, Madersbacher S, Djavan B, Schatzl G, Marberger M, Schmidbauer CP. Morbidity of the evaluation of the lower urinary tract with transurethral multichannel pressure-flow studies. *J Urol*. 1998;159(1):191-194.

76. Belal M, Abrams P. Noninvasive methods of diagnosing bladder outlet obstruction in men. Part 1: Nonurodynamic approach. *J Urol*. 2006;176(1):22-28.

77. Rademakers KLJ, van Koeveringe GA, Oelke M, FORCE Research Group, Maastricht and Hannover. Ultrasound detrusor wall thickness measurement in combination with

bladder capacity can safely detect detrusor underactivity in adult men. *World J Urol*. July 2016.

78. McIntosh SL, Drinnan MJ, Griffiths CJ, Robson WA, Ramsden PD, Pickard RS. Noninvasive assessment of bladder contractility in men. *J Urol*. 2004;172(4 Pt 1):1394-1398.

79. Koeveringe GA van, Rademakers KLJ. Non-invasive Diagnostics for Detrusor Underactivity/Underactive Bladder. In: *Underactive Bladder*. Springer, Cham; 2017:13-16.

80. Hillary C, Osman N, Chapple C. Indications for treatment. In: Chapple C, Wein A, Osman N, eds. *Underactive Bladder*. 1st Edition. Switzerland: Springer; 2017:31-39.

81. Ghalayini IF, Al-Ghazo MA, Pickard RS. A prospective randomized trial comparing transurethral prostatic resection and clean intermittent self-catheterization in men with chronic urinary retention. *BJU Int*. 2005;96(1):93-97.

82. Thomas A w., Cannon A, Bartlett E, Ellis-Jones J, Abrams P. The natural history of lower urinary tract dysfunction in men: the influence of detrusor underactivity on the outcome after transurethral resection of the prostate with a minimum 10-year urodynamic follow-up. *BJU Int*. 2004;93(6):745-750.

83. Jhang J-F, Jiang Y-H, Lee C-L, Kuo H-C. Long-term follow up and predictive factors for successful outcome of transurethral incision of the bladder neck in women with detrusor underactivity. *J Formos Med Assoc Taiwan Yi Zhi*. September 2015.

84. Osman NI, Chapple CR, Wein AJ. Reflections and the Way Forward. In: *Underactive Bladder*. Springer, Cham; 2017:83-85.

85. Andersson KE. Pharmacological Treatment of Underactive Bladder. In: *Underactive Bladder*. Springer, Cham; 2017:39-48.

86. Barendrecht MM, Oelke M, Laguna MP, Michel MC. Is the use of parasympathomimetics for treating an underactive urinary bladder evidence-based? *BJU Int*. 2007;99(4):749-752.

87. Guang-Jun D, Feng-Bin G, Xun-Bo J. A_1-blockers in the management of acute urinary retention secondary to benign prostatic hyperplasia: a systematic review and meta-analysis. *Ir J Med Sci*. 2015;184(1):23-30.

88. Yamanishi T, Yasuda K, Homma Y, Kawabe K, Morita T. A Multicenter Placebo-Controlled, Double-Blind Trial of Urapidil, an α-Blocker, on Neurogenic Bladder Dysfunction. *Eur Urol*. 1999;35(1):45-51.

89. Sekido N, Kida J, Mashimo H, Wakamatsu D, Okada H, Matsuya H. Promising Effects of a Novel EP2 and EP3 Receptor Dual Agonist, ONO-8055, on Neurogenic Underactive Bladder in a Rat Lumbar Canal Stenosis Model. *J Urol*. 2016;196(2):609-616.

90. Drossaerts J, Jairam R, Koeveringe GA van, Kerrebroeck PEV van. Neurostimulation and Neuromodulation for the Treatment for the Underactive Bladder. In: *Underactive Bladder*. Springer, Cham; 2017:57-61.

91. van Kerrebroeck PEV, van Voskuilen AC, Heesakkers JPFA, et al. Results of Sacral Neuromodulation Therapy for Urinary Voiding Dysfunction: Outcomes of a Prospective, Worldwide Clinical Study. *J Urol*. 2007;178(5):2029-2034.

92. Bright E, Cotterill N, Drake M, Abrams P. Developing and validating the International Consultation on Incontinence Questionnaire bladder diary. *Eur Urol*. 2014;66(2):294-300.

93. Madsen PO, Iversen P. A Point System for Selecting Operative Candidates. In: Jr FH, F.A.C.S SBMD JD, eds. *Benign Prostatic Hypertrophy*. Springer New York; 1983:763-765.

94. Shaw C, Matthews RJ, Perry SI, et al. Validity and reliability of a questionnaire to measure the impact of lower urinary tract symptoms on quality of life: the Leicester Impact Scale. *Neurourol Urodyn*. 2004;23(3):229-236.

95. Rodríguez LV, Blander DS, Dorey F, Raz S, Zimmern P. Discrepancy in patient and physician perception of patient's quality of life related to urinary symptoms. *Urology*. 2003;62(1):49-53.

96. Hansen BJ, Mortensen S, Mensink HJ, et al. Comparison of the Danish Prostatic Symptom Score with the International Prostatic Symptom Score, the Madsen-Iversen and Boyarsky symptom indexes. ALFECH Study Group. *Br J Urol*. 1998;81(1):36-41.

97. Castro Diaz D, Robinson D, Bosch R, et al. Patient-Reported Outcome Assessment. In: Abrams P, Cardozo L, Wagg A, Wein A, eds. *Incontinence, 6th Edition 2017: International Consultation on Incontinence, Tokyo, September 2016*. International Continence Society; 2017:541-599.

98. Deshpande PR, Rajan S, Sudeepthi BL, Abdul Nazir CP. Patient-reported outcomes: A new era in clinical research. *Perspect Clin Res*. 2011;2(4):137-144.

99. FDA. Guidance for industry: patient-reported outcome measures: use in medical product development to support labeling claims. December 2009.

100. Weldring T, Smith SMS. Patient-Reported Outcomes (PROs) and Patient-Reported Outcome Measures (PROMs). *Health Serv Insights*. 2013;6:61-68.

101. Bowling A. *Measuring Health: A Review of Quality of Life Measurement Scales*. 3 edition. Maidenhead, Berkshire, England; New York, NY: Open University Press; 2004.

102. Frost MH, Reeve BB, Liepa AM, Stauffer JW, Hays RD. What Is Sufficient Evidence for the Reliability and Validity of Patient-Reported Outcome Measures? *Value Health*. 2007;10:S94-S105.

103. Oppenheim AN. *Questionnaire Design, Interviewing and Attitude Measurement*. London: Bloomsbury Publishing; 2000.

104. Streiner DL, Norman GR, Cairney J. *Health Measurement Scales: A Practical Guide to Their Development and Use*. Third edition. Oxford University Press; 2004.

105. Kingsley C, Patel S. Patient-reported outcome measures and patient-reported experience measures. *BJA Educ*. 2017;17(4):137-144.

106. Karmakar D, Mostafa A, Abdel-Fattah M. A new validated score for detecting patient-reported success on postoperative ICIQ-SF: a novel two-stage analysis from two large RCT cohorts. *Int Urogynecology J*. 2017;28(1):95-100.

107. Twiss C, Triaca V, Anger J, et al. Validating the Incontinence Symptom Severity Index: A Self-Assessment Instrument for Voiding Symptom Severity in Women. *J Urol*. 2009;182(5):2384-2391.

108. Hays RD, Anderson R, Revicki D. Psychometric considerations in evaluating health-related quality of life measures. *Qual Life Res Int J Qual Life Asp Treat Care Rehabil*. 1993;2(6):441-449.

109. Stewart A. Psychometric considerations in functional status instruments. In: Lipkin M, ed. *Functional Status Measurement in Primary Care*. New York: Springer; 1990:3-26.

110. Wynd CA, Schmidt B, Schaefer MA. Two quantitative approaches for estimating content validity. *West J Nurs Res*. 2003;25(5):508-518.

111. Coyne K, Sexton C. Patient-reported outcomes: From development to utilization. In: Cardozo L, Staskin D, eds. *Textbook of Female Urology and Urogynecology Fourth Edition*. Vol Volume one. Fourth Edition. CRC Press; 2017:100-108.

112. Patrick DL, Burke LB, Gwaltney CJ, et al. Content Validity—Establishing and Reporting the Evidence in Newly Developed Patient-Reported Outcomes (PRO) Instruments for Medical Product Evaluation: ISPOR PRO Good Research Practices Task Force Report: Part 1—Eliciting Concepts for a New PRO Instrument. *Value Health*. 2011;14(8):967-977.

113. Litwin MS. *How to Measure Survey Reliability and Validity*. SAGE Publications; 1995.

114. FDA. Guidance for Industry and FDA Staff: Qualification for Drug Development Tools. January 2014.

115. Patrick DL, Burke LB, Gwaltney CJ, et al. Content Validity—Establishing and Reporting the Evidence in Newly Developed Patient-Reported Outcomes (PRO) Instruments for Medical Product Evaluation: ISPOR PRO Good Research Practices Task Force Report: Part 2—Assessing Respondent Understanding. *Value Health*. 2011;14(8):978-988.

116. Lasch KE, Marquis P, Vigneux M, et al. PRO development: rigorous qualitative research as the crucial foundation. *Qual Life Res Int J Qual Life Asp Treat Care Rehabil*. 2010;19(8):1087-1096.

117. Leidy NK, Vernon M. Perspectives on patient-reported outcomes : content validity and qualitative research in a changing clinical trial environment. *PharmacoEconomics*. 2008;26(5):363-370.

118. Lasch KE, Jernigan KA, Scott JA, Piault-Louis E. Interviews with cancer fatigue patients. *J Clin Oncol*. 2009;27(15S):e17577-e17577.

119. Morse JM. The Significance of Saturation. *Qual Health Res*. 1995;5(2):147-149.

120. Rothman M, Burke L, Erickson P, Leidy NK, Patrick DL, Petrie CD. Use of existing patient-reported outcome (PRO) instruments and their modification: the ISPOR Good Research

Practices for Evaluating and Documenting Content Validity for the Use of Existing Instruments and Their Modification PRO Task Force Report. *Value Health J Int Soc Pharmacoeconomics Outcomes Res*. 2009;12(8):1075-1083.

121. Kerr C, Nixon A, Wild D. Assessing and demonstrating data saturation in qualitative inquiry supporting patient-reported outcomes research. *Expert Rev Pharmacoecon Outcomes Res*. 2010;10(3):269-281.

122. Guest G, Bunce A, Johnson L. How Many Interviews Are Enough?: An Experiment with Data Saturation and Variability. *Field Methods*. 2006;18(1):59-82.

123. Sandelowski M. Sample size in qualitative research. *Res Nurs Health*. 1995;18(2):179-183.

124. Cotterill N, Norton C, Avery KNL, Abrams P, Donovan JL. A Patient-Centered Approach to Developing a Comprehensive Symptom and Quality of Life Assessment of Anal Incontinence: *Dis Colon Rectum*. 2008;51(1):82-87.

125. Leidy NK, Wilcox TK, Jones PW, et al. Development of the EXAcerbations of Chronic Obstructive Pulmonary Disease Tool (EXACT): a patient-reported outcome (PRO) measure. *Value Health J Int Soc Pharmacoeconomics Outcomes Res*. 2010;13(8):965-975.

126. Braun V, Clarke V. Using thematic analysis in psychology. *Qual Res Psychol*. 2006;3(2):77-101.

127. Silverman D. *Doing Qualitative Research*. 3 edition. London ; Thousand Oaks, Calif: SAGE Publications Ltd; 2009.

128. Holloway I, Todres L. The Status of Method: Flexibility, Consistency and Coherence. *Qual Res*. 2003;3(3):345-357.

129. Smith J, Firth J. Qualitative data analysis: the framework approach. *Nurse Res*. 2011;18(2):52-62.

130. Starks H, Trinidad SB. Choose Your Method: A Comparison of Phenomenology, Discourse Analysis, and Grounded Theory. *Qual Health Res*. 2007;17(10):1372-1380.

131. Strauss A, Corbin J. Grounded theory methodology: an overview. In: *Handbook of Qualitative Research*. CA, USA: SAGE Publicaations Ltd.; 1994:273-285.

132. Hsieh H-F, Shannon SE. Three approaches to qualitative content analysis. *Qual Health Res*. 2005;15(9):1277-1288.

133. Glaser B, Strauss A. The constant comparative methods of qualitative analysis. In: *The Discovery of Grounded Theory: Strategies for Qualitative Research*. Chicago, IL: Aldine Publishing Co; 1967:101-106.

134. Thomas DR. A General Inductive Approach for Analyzing Qualitative Evaluation Data. *Am J Eval*. 2006;27(2):237-246.

135. Pope C, Ziebland S, Mays N. Analysing qualitative data. *BMJ*. 2000;320(7227):114-116.

136. Muhr T. *Users's Manual for Atlas.Ti 5.0.* Berlin: Scientific Software Development; 2004.

137. NVivo. *NVivo Qualitative Data Analysis Software*. QSR International Pty Ltd; 2012.

138. Patton MQ. *Qualitative Research & Evaluation Methods*. 3 edition. Thousand Oaks, Calif: SAGE Publications, Inc; 2002.

139. Campbell JL, Quincy C, Osserman J, Pedersen OK. Coding In-depth Semistructured Interviews Problems of Unitization and Intercoder Reliability and Agreement. *Sociol Methods Res*. 2013;42(3):294-320.

140. Whittemore R, Chase SK, Mandle CL. Validity in qualitative research. *Qual Health Res*. 2001;11(4):522-537.

141. Willis GB. *Cognitive Interviewing: A Tool for Improving Questionnaire Design*. Thousand oaks, CA: SAGE Publications, Inc; 2005.

142. Peat J. *Health Science Research: A Handbook of Quantitative Methods*. 1 edition. London ; Thousand Oaks: SAGE Publications Ltd; 2001.

143. Stone A, Shiffman S, Atienza A, Nebeling L. *The Science of Real-Time Data Capture: Self-Reports in Health Research*. Oxford, New York: Oxford University Press; 2007.

144. Hambleton RK, Swaminathan H, Rogers H. *Fundamentals of Item Response Theory*. 1 edition. Newbury Park, Calif: SAGE Publications, Inc; 1991.

145. Mallen CD, Dunn KM, Thomas E, Peat G. Thicker paper and larger font increased response and completeness in a postal survey. *J Clin Epidemiol*. 2008;61(12):1296-1300.

146. Coons SJ, Gwaltney CJ, Hays RD, et al. Recommendations on Evidence Needed to Support Measurement Equivalence between Electronic and Paper-Based Patient-Reported Outcome (PRO) Measures: ISPOR ePRO Good Research Practices Task Force Report. *Value Health*. 2009;12(4):419-429.

147. Tourangeau R. Cognitive science and survey methods. In: Jabine T, Straf M, Tanur J, eds. *Cognitive Aspects of Survey Design: Building a Bridge Between Disciplines*. Washington DC: National Academy Press; 1984:73-79.

148. Kucan L, Beck IL. Thinking Aloud and Reading Comprehension Research: Inquiry, Instruction, and Social Interaction. *Rev Educ Res*. 1997;67(3):271-299.

149. Abraham L, Hareendran A, Mills IW, et al. Development and validation of a quality-of-life measure for men with nocturia. *Urology*. 2004;63(3):481-486.

150. Jackson S, Donovan J, Brookes S, Eckford S, Swithinbank L, Abrams P. The Bristol female lower urinary tract symptoms questionnaire: development and psychometric testing. *Br J Urol*. 1996;77(6):805–812.

151. O'Mahony PG, Rodgers H, Thomson RG, Dobson R, James OFW. Is the SF-36 suitable for assessing health status of older stroke patients? *Age Ageing*. 1998;27(1):19-22.

152. Fries J, Rose M, Krishnan E. The promise of better outcome assessment: Responiveness, floor and ceiling effects, and internet administration. *J Rheumatol*. 2011;38(8):1759-64.

153. Nitti VW. The Prevalence of Urinary Incontinence. *Rev Urol*. 2001;3(Suppl 1):S2-S6.

154. Nunnally J, Bernstein I. The assessment of reliability. In: *Psychometric Theory*. Third ed. McGraw-Hill Inc; 1994:248-292.

155. Thomas S, Nay R, Moore K, et al. *Continent Outcomes Measurement Suite (Final Report)*. Australian Government Department of Health and Aging; 2006.

156. Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. *Psychol Bull*. 1979;86(2):420-428.

157. Landis JR, Koch GG. The Measurement of Observer Agreement for Categorical Data. *Biometrics*. 1977;33(1):159.

158. Hays RD, Hadorn D. Responsiveness to Change: An Aspect of Validity, Not a Separate Dimension. *Qual Life Res*. 1992;1(1):73-75.

159. Patrick DL, Chiang YP. Measurement of health outcomes in treatment effectiveness evaluations: conceptual and methodological challenges. *Med Care*. 2000;38(9 Suppl):II14-25.

160. Streiner DL, Norman GR, Cairney J. Measuring change. In: *Health Measurement Scales: A Practical Guide to Their Development and Use*. Third edition. Oxford University Press; 2004:194-227.

161. Guyatt G, Walter S, Norman G. Measuring change over time: assessing the usefulness of evaluative instruments. *J Chronic Dis*. 1987;40(2):171-178.

162. Kirshner B, Guyatt G. A methodological framework for assessing health indices. *J Chronic Dis*. 1985;38(1):27-36.

163. Brookes ST, Donovan JL, Wright M, Jackson S, Abrams P. A scored form of the Bristol Female Lower Urinary Tract Symptoms questionnaire: Data from a randomized controlled trial of surgery for women with stress incontinence. *Am J Obstet Gynecol*. 2004;191(1):73-82.

164. Yong AG, Pearce S. A beginner's guide to factor analysis: Focusing on exploratory factor analysis. *Tutor Quant Methods Psychol*. 2013;9(2):79–94.

165. McDonald RP. *Factor Analysis and Related Methods*. Psychology Press; 1985.

166. Petrillo J, Cano SJ, McLeod LD, Coon CD. Using Classical Test Theory, Item Response Theory, and Rasch Measurement Theory to Evaluate Patient-Reported Outcome Measures: A Comparison of Worked Examples. *Value Health*. 2015;18(1):25-34.

167. Novick MR. The axioms and principal results of classical test theory. *J Math Psychol*. 1966;3(1):1-18.

168. Lord FM, Novick MR. *Statistical Theories of Mental Test Scores*. Reading: Addison-Welsley Publishing Company; 1968.

169. Cappelleri JC, Jason Lundy J, Hays RD. Overview of Classical Test Theory and Item Response Theory for the Quantitative Assessment of Items in Developing Patient-Reported Outcomes Measures. *Clin Ther*. 2014;36(5):648-662.

170. Flynn KE, Dombeck CB, DeWitt EM, Schulman KA, Weinfurt KP. Using Item Banks to Construct Measures of Patient Reported Outcomes in Clinical Trials: Investigator Perceptions. *Clin Trials Lond Engl*. 2008;5(6):575-586.

171. Rasch G. *Probabilistic Models for Some Intelligence and Attainment Tests*. Expanded ed edition. Chicago: University of Chicago Press; 1980.

172. Bjorner JB, Kosinski M, Ware JE. The feasibility of applying item response theory to measures of migraine impact: a re-analysis of three clinical studies. *Qual Life Res Int J Qual Life Asp Treat Care Rehabil*. 2003;12(8):887-902.

173. Becker J, Schwartz C, Saris-Baglama RN, Kosinski M, Bjorner JB. Using Item Response Theory (IRT) for Developing and Evaluating the Pain Impact Questionnaire (PIQ-6^{TM}). *Pain Med*. 2007;8(suppl_3):S129-S144.

174. Acquadro C, Conway K, Giroudet C, Mear I. *Linguistic Validation Manual for Health Outcome Assessments*. Second editoin. Lyon, France: Mapi Institute; 2004.

175. Wild D, Grove A, Martin M, et al. Principles of Good Practice for the Translation and Cultural Adaptation Process for Patient-Reported Outcomes (PRO) Measures: Report of the ISPOR Task Force for Translation and Cultural Adaptation. *Value Health*. 2005;8(2):94-104.

176. Barry MJ, Fowler FJ, O'Leary MP, et al. The American Urological Association symptom index for benign prostatic hyperplasia. The Measurement Committee of the American Urological Association. *J Urol*. 1992;148(5):1549-1557; discussion 1564.

177. Blaivas JG, Tsui JF, Mekel G, et al. Validation of the lower urinary tract symptom score. *Can J Urol*. 2015;22(5):7952-7958.

178. Brandt BA, Angün Ç, Coyne KS, Doshi S, Bavendam T, Kopp ZS. LUTS patient reported outcomes tool: Linguistic validation in 10 European languages. *Neurourol Urodyn*. 2013;32(1):75-81.

179. Underactive Bladder Foundation. UAB-q. *Underactive Bladder Found*. September 2016.

180. Basra R, Artibani W, Cardozo L, et al. Design and Validation of a New Screening Instrument for Lower Urinary Tract Dysfunction: The Bladder Control Self-Assessment Questionnaire (B-SAQ). *Eur Urol*. 2007;52(1):230-238.

181. Sahai A, Dowson C, Cortes E, et al. Validation of the bladder control self-assessment questionnaire (B-SAQ) in men: Validation of the B-SAQ in men. *BJU Int*. 2014;113(5):783-788.

182. Blaivas JG, Panagopoulos G, Weiss JP, Somaroo C. Validation of the Overactive Bladder Symptom Score. *J Urol*. 2007;178(2):543-547.

183. Meyhoff HH, Hald T, Nordling J, Andersen JT, Bilde T, Walter S. A new patient weighted symptom score system (DAN-PSS-1). Clinical assessment of indications and outcomes of

transurethral prostatectomy for uncomplicated benign prostatic hyperplasia. *Scand J Urol Nephrol*. 1993;27(4):493-499.

184. Hansen BJ, Flyger H, Brasso K, et al. Validation of the self-administered Danish Prostatic Symptom Score (DAN-PSS-1) system for use in benign prostatic hyperplasia. *Br J Urol*. 1995;76(4):451-458.

185. Donovan JL, Abrams P, Peters TJ, et al. The ICS-'BPH'Study: the psychometric validity and reliability of the ICSmale questionnaire. *Br J Urol*. 1996;77:554–562.

186. Donovan JL, Peters TJ, Abrams P, Brookes ST, De La Rosette J, Schäfer W. Scoring the short form ICSmaleSF questionnaire. *J Urol*. 2000;164(6):1948–1955.

187. Coyne KS, Zyczynski T, Margolis MK, Elinoff V, Roberts RG. Validation of an overactive bladder awareness tool for use in primary care settings. *Adv Ther*. 2005;22(4):381–394.

188. Coyne K, Revicki D, Hunt T, et al. Psychometric validation of an overactive bladder symptom and health-related quality of life questionnaire: the OAB-q. *Qual Life Res*. 2002;11(6):563–574.

189. Coyne KS, Gelhorn H, Thompson C, Kopp ZS, Guan Z. The psychometric validation of a 1-week recall period for the OAB-q. *Int Urogynecology J*. 2011;22(12):1555-1563.

190. Donovan JL, Kay HE, Peters TJ, et al. Using the ICSOoL to measure the impact of lower urinary tract symptoms on quality of life: evidence from the ICS-'BPH' Study. International Continence Society--Benign Prostatic Hyperplasia. *Br J Urol*. 1997;80(5):712-721.

191. Kelleher CJ, Cardozo LD, Khullar V, Salvatore S. A new questionnaire to assess the quality of life of urinary incontinent women. *BJOG Int J Obstet Gynaecol*. 1997;104(12):1374–1379.

192. Margolis MK, Vats V, Coyne KS, Kelleher C. Establishing the content validity of the King's Health Questionnaire in men and women with overactive bladder in the US. *The Patient*. 2011;4(3):177-187.

193. Kelleher CJ, Pleil AM, Reese PR, Burgess SM, Brodish PH. How much is enough and who says so? *BJOG Int J Obstet Gynaecol*. 2004;111(6):605-612.

194. Reese PR, Pleil AM, Okano GJ, Kelleher CJ. Multinational study of reliability and validity of the King's Health Questionnaire in patients with overactive bladder. *Qual Life Res Int J Qual Life Asp Treat Care Rehabil*. 2003;12(4):427-442.

195. Patrick DL, Burke LB, Powers JH, et al. Patient-reported outcomes to support medical product labeling claims: FDA perspective. *Value Health J Int Soc Pharmacoeconomics Outcomes Res*. 2007;10 Suppl 2:S125-137.

196. Coyne K, Kelleher C. Patient reported outcomes: The ICIQ and the state of the art. *Neurourol Urodyn*. 2010;29(4):645-651.

197. Donovan J l., Kay H e., Peters T j., et al. Using the ICSQoL to measure the impact of lower urinary tract symptoms on quality of life: evidence from the ICS–'BPH' study. *Br J Urol*. 1997;80(5):712-721.

198. Coyne KS, Barsdorf AI, Thompson C, et al. Moving towards a comprehensive assessment of lower urinary tract symptoms (LUTS). *Neurourol Urodyn*. 2012;31(4):448-454.

199. Abrams P, Avery K, Gardener N, Donovan J, ICIQ Advisory Board. The International Consultation on Incontinence Modular Questionnaire: www.iciq.net. *J Urol*. 2006;175(3 Pt 1):1063-1066; discussion 1066.

200. Abrams P, Cardozo L, Fall M, et al. The standardisation of terminology of lower urinary tract function: report from the Standardisation Sub-committee of the International Continence Society. *Neurourol Urodyn*. 2002;21(2):167-178.

201. Haylen BT, Maher CF, Barber MD, et al. An International Urogynecological Association (IUGA) / International Continence Society (ICS) Joint Report on the Terminology for Female Pelvic Organ Prolapse (POP): International Urogynecological Association (IUGA). *Neurourol Urodyn*. 2016;35(2):137-168.

202. Abrams P, Cardozo L, Fall M, et al. The standardisation of terminology in lower urinary tract function: report from the standardisation sub-committee of the International Continence Society. *Urology*. 2003;61(1):37-49.

203. Cho MC, Ha SB, Park J, et al. Impact of Detrusor Underactivity on Surgical Outcomes of Laser Prostatectomy: Comparison in Serial 12-Month Follow-Up Outcomes Between Potassium-Titanyl-Phosphate Photoselective Vaporization of the Prostate (PVP) and Holmium Laser Enucleation of the Prostate (HoLEP). *Urology*. 2016;91:158-166.

204. Thomas AW, Cannon A, Bartlett E, Ellis-Jones J, Abrams P. The natural history of lower urinary tract dysfunction in men: minimum 10-year urodynamic followup of transurethral resection of prostate for bladder outlet obstruction. *J Urol*. 2005;174(5):1887-1891.

205. Morrow SL. Quality and trustworthiness in qualitative research in counseling psychology. *J Couns Psychol*. 2005;52(2):250-260.

206. Hertzog MA. Considerations in determining sample size for pilot studies. *Res Nurs Health*. 2008;31(2):180-191.

207. Chuang Y-C, Plata M, Lamb LE, Chancellor MB. Underactive Bladder in Older Adults. *Clin Geriatr Med*. 2015;31(4):523-533.

208. Chung A, Noguchi N, Chan L, Tse V. Voiding dysfunction in older men. *Curr Opin Urol*. 2016;26(2):177-83.

209. Teunissen T a. M, Lagro-Janssen ALM, Bosch WJHM van den, Hoogen HJM van den. Prevalence of urinary, fecal and double incontinence in the elderly living at home. *Int Urogynecology J*. 2003;15(1):10-13.

210. Thomas TM, Plymat KR, Blannin J, Meade TW. Prevalence of urinary incontinence. *Br Med J*. 1980;281(6250):1243-1245.

211. Nnabugwu II, Udeh EI, Enivwenae OA, Ugwumba FO, Ozoemena OF. Reducing the burden of regular indwelling urinary catheter changes in the catheter clinics: the opinion of patients and relatives on the practice of self-catheterization. *Patient Prefer Adherence*. 2014;8:1179-1183.

212. Altman DG. *Practical Statistics for Medical Research*. CRC Press; 1990.

213. Geisser ME, Clauw DJ, Strand V, Gendreau RM, Palmer R, Williams DA. Contributions of change in clinical status parameters to Patient Global Impression of Change (PGIC) scores among persons with fibromyalgia treated with milnacipran. *Pain*. 2010;149(2):373-378.

214. Kinchen KS, Obenchain R, Swindle R. Impact of duloxetine on quality of life for women with symptoms of urinary incontinence. *Int Urogynecol J Pelvic Floor Dysfunct*. 2005;16(5):337-344.

215. Wyndaele J. Complications of intermittent catheterization: their prevention and treatment. *Publ Online 18 Sept 2002 Doi101038sjsc3101348*. 2002;40(10):536-41.

216. Crawford B, Bouhassira D, Wong A, Dukes E. Conceptual adequacy of the neuropathic pain symptom inventory in six countries. *Health Qual Life Outcomes*. 2008;6:62.

217. Hugosson J, Stranne J, Carlsson SV. Radical retropubic prostatectomy: A review of outcomes and side-effects. *Acta Oncol*. 2011;50(sup1):92-97.

218. Hanno P, Lin A, Nordling J, et al. Bladder pain syndrome international consultation on incontinence. *Neurourol Urodyn*. 2010;29(1):191-198.

219. Reeve BB, Wyrwich KW, Wu AW, et al. ISOQOL recommends minimum standards for patient-reported outcome measures used in patient-centered outcomes and comparative effectiveness research. *Qual Life Res*. 2013;22(8):1889-1905.

220. Limperopoulos C, Majnemer A, Steinbach CL, Shevell M. Equivalence Reliability of the Vineland Adaptive Behavior Scale Between In-Person and Telephone Administration. *Phys Occup Ther Pediatr*. 2006;26(1):115-127.

221. Bushnell DM, Reilly MC, Galani C, et al. Validation of Electronic Data Capture of the Irritable Bowel Syndrome—Quality of Life Measure, the Work Productivity and Activity Impairment Questionnaire for Irritable Bowel Syndrome and the EuroQol. *Value Health*. 2006;9(2):98-105.

222. Stone AA, Shiffman S, Schwartz JE, Broderick JE, Hufford MR. Patient non-compliance with paper diaries. *Bmj*. 2002;324(7347):1193–1194.

223. Velikova G, Wright EP, Smith AB, et al. Automated collection of quality-of-life data: a comparison of paper and computer touch-screen questionnaires. *J Clin Oncol*. 1999;17(3):998–998.

224. Aiello EJ, Taplin S, Reid R, et al. In a randomized controlled trial, patients preferred electronic data collection of breast cancer risk-factor information in a mammography setting. *J Clin Epidemiol*. 2006;59(1):77-81.

225. Bischoff-Ferrari HA. Validation and patient acceptance of a computer touch screen version of the WOMAC 3.1 osteoarthritis index. *Ann Rheum Dis*. 2005;64(1):80-84.

226. Hess R, Santucci A, McTigue K, Fischer G, Kapoor W. Patient Difficulty Using Tablet Computers to Screen in Primary Care. *J Gen Intern Med*. 2008;23(4):476-480.

227. Bellamy N, Campbell J, Hill J, Band P. A comparative study of telephone versus onsite completion of the WOMAC 3.0 osteoarthritis index. *J Rheumatol*. 2002;29(4):783–786.

228. Lungenhausen M, Lange S, Maier C, Schaub C, Trampisch HJ, Endres HG. Randomised controlled comparison of the Health Survey Short Form (SF-12) and the Graded Chronic Pain Scale (GCPS) in telephone interviews versus self-administered questionnaires. Are the results equivalent? *BMC Med Res Methodol*. 2007;7(1):50.

229. Resnik LJ, Clark MA, Borgia M. Telephone and face to face methods of assessment of veteran's community reintegration yield equivalent results. *BMC Med Res Methodol*. 2011;11(1):98.

230. Gwaltney CJ, Shields AL, Shiffman S. Equivalence of Electronic and Paper-and-Pencil Administration of Patient-Reported Outcome Measures: A Meta-Analytic Review. *Value Health*. 2008;11(2):322-333.

231. Schulenberg SE, Yutrzenka BA. The equivalence of computerized and paper-and-pencil psychological instruments: Implications for measures of negative affect. *Behav Res Methods Instrum Comput*. 1999;31(2):315-321.

232. Juniper EF, Langlands JM, Juniper BA. Patients may respond differently to paper and electronic versions of the same questionnaires. *Respir Med*. 2009;103(6):932-934.

233. Uren AD, Cotterill N, Parke SE, Abrams P. Psychometric equivalence of electronic and telephone completion of the ICIQ modules. *Neurourol Urodyn*. 2017;36(5):1342-1349.

234. Coons SJ, Gwaltney CJ, Hays RD, et al. Recommendations on Evidence Needed to Support Measurement Equivalence between Electronic and Paper-Based Patient-Reported Outcome (PRO) Measures: ISPOR ePRO Good Research Practices Task Force Report. *Value Health*. 2009;12(4):419-429.

235. Bjorner JB, Kosinski M, Jr JEW. Calibration of an item pool for assessing the burden of headaches: An application of item response theory to the Headache Impact Test (HIT^{TM}). *Qual Life Res*. 2003;12(8):913-933.

236. Gammie A, Kaper M, Steup A, et al. Signs and symptoms that distinguish detrusor underactivity from mixed detrusor underactivity and bladder outlet obstruction in male patients. *Neurourol Urodyn*. 2018;37(4):1501-1505.

Appendix 1 Clinical review panel

Name	Role	Location
Ms Debbie Delgado	Urology Research Nurse	Southmead Hospital, Bristol, UK
Dr Michael Drinnan	Head of Clinical Engineering with the Regional Medical Physics Department	Freeman Hospital, Newcastle upon Tyne, UK
Mr Jason Britton	Clinical Scientist	Leeds NHS teaching Hospital, UK
Prof Christopher Chapple	Urological Surgeon	Sheffield Teaching Hospital, UK
Dr Nadir Osman	Clinical Research Fellow	University of Sheffield and Royal Hallamshire Hospital, UK
Prof Robert Pickard	Consultant Urologist	Freeman Hospital, Newcastle Upon Tyne, UK
Prof Philip Smith	Assistant Professor of Surgery	Centre for Continence and Voiding Disorders, Connecticut, US
Dr Hashim Hashim	Consultant Urologist	Southmead Hospital, Bristol, UK
Prof Marcus Drake	Consultant Urologist and Professor at the University of Bristol	Southmead Hospital, Bristol, UK

Appendix 2 Patient interview information sheet

ICIQ Study

Developing a Patient Reported Outcome measure for individuals with underactive bladder

Introduction

We would like to invite you to help us develop a new questionnaire to assess underactive bladder symptoms and their impact on quality of life. The questionnaire will then be available for use by patients to measure the effectiveness of treatments for underactive bladder symptoms.

The questionnaire is called the ICIQ (International Consultation on Incontinence Questionnaire). At the moment, the questionnaire is being developed for people suffering from an underactive bladder, and we are asking for patients to help us by giving us their opinion on the content of the questionnaire with the aim of developing some new questions. This will help us to improve the questionnaire and make it a more useful tool in assessing underactive bladder symptoms.

What will I have to do if I take part?

The study will involve 30 patients. If you would like to participate in the study, all that you would be required to do is complete a consent form and discuss your experiences of underactive bladder with a member of our research team – xxxxxx.

The discussion will take a maximum of 60 minutes and will be held either in a private room within the Outpatients Clinic (at the Urology Dept.), at your own home, or over the phone, depending on what works best for you. Once the interview has finished, your participation in the study is completed.

Are there any possible benefits?

The information that we get from this study will help us to develop a questionnaire schedule for measuring patients' underactive bladder symptoms and their impact. The questionnaire will then be used to measure the effectiveness of treatments for these symptoms.

Do I have to take part?

You do not have to take part and your participation is entirely voluntary. If you prefer not to take part you do not have to give a reason and your future treatment will not be affected. If you do decide to take part, but later change your mind, you can withdraw at any time without affecting your health care.

Compensation

As a small thank you for giving up your time to participate we would like to give you a £50 voucher that is redeemable at a number of high street stores. If at any point during the interview you decide to withdraw your participation you will still receive the voucher.

What do I do now?

You will be asked by the researcher named above to complete a consent form and participate in a discussion about your experiences of underactive bladder. Once the discussion is finished, your participation in the study is complete.

Please note that the discussion is *strictly confidential and anonymous*. Information gathered will only be used for the purposes of the research mentioned above and the results presented such that the information from a single individual cannot be identified. Access to the information you provide will be limited to members of the research team and will not be seen by the doctors or nurses in the clinic, or by your GP.

Thank you very much for considering taking part in our research.

Appendix 3 Concept elicitation interview schedule

Assessing the signs and symptoms of Underactive Bladder
Discussion Guide for Interviews

OPENING (5 MINUTES)

- Introduce interviewer
- Remind participants of the purpose and format of the interview

" During this interview, we would like to discuss with you the signs and symptoms as well as the impact of Underactive Bladder/Bladder Outlet obstruction on your life. We are developing a questionnaire that will reflect the views of people like you to help us better understand the effect of potential treatments on people's lives. So we want to hear about your experiences with UAB; we will start with the symptoms you've experienced, and the treatment you have received. Then we will explore how your condition has affected your life."

  - Confidentiality:

    "As researchers we would like to assure you that all of the information that you share with me today will be treated as confidential by the Bristol Urological Institute. We will type up transcripts from the recordings but any names or other personal identifiable information will be replaced with pseudonyms. We will use the transcripts for our results, but it will be reported in such a way that no individuals can be identified."

HISTORY OF Underactive Bladder (UAB) (10 MINUTES)

Before we begin, could you list any other illnesses you currently have, besides your bladder condition?

How would you describe your bladder condition in one sentence?

Could you describe when your bladder issues first appeared?

- What made you first see a health professional?
    - When was this?
    - How long were you experiencing these symptoms for?
- Have you talked to anyone else about your condition?
- How have you managed the condition?
- Have you had any difficulties with accessing treatment?

What do you know about the term underactive bladder/detrusor underactivity?

Signs and symptoms of UAB (15 MINUTES)
Discussion of storage and voiding symptoms

We're now going to talk for a bit about your bladder and how it is for you when it is filling and when you actually go to the toilet.

During the time in between toilet visits, do you experience any sensations?

- If they mention leakage, do they feel it?
- What does their bladder feel like? E.g. prickling, tingling.

How does it feel as your bladder is filling up, shortly before you visit the toilet?

- Feelings of urgency – what brings on this feeling?

How soon do you visit the toilet after you first need to go?

- Do they feel the need to rush to the toilet?

If you have to wait to use the toilet, how does it feel?

Could you describe how it feels when you go to pass urine?

- Difficulty passing urine
- Straining or "letting go" – how do they strain? Hands, muscles etc
- Is anything different when you go to toilet when other people are present?
- Intermittency – number of times per urination, how often.
- Splitting or spraying of the flow
- Any dribbling – at what point of the urination does this occur?
- Any discomfort?

Do you feel satisfied or that you have fully emptied your bladder afterwards?

- Describe these feelings, or lack of feelings.
- Where would the understanding of incomplete emptying come from.
- Or, how do they know they have properly emptied.

Other prompts:

- Number of daytime visits to the toilet
- Waking up at night to urinate

What do you think is the cause of these symptoms?

What symptom bothers you the most?

Are there any concerns you have about your symptoms?

Have you experienced any difficulties with your bowels recently?

Have you had any UTIs over the last 12 months?

Do you have to use a catheter?

These questions are intended to clarify the participant's earlier statements, and shed light on any symptoms that they may have missed in the discussion so far. Each question **must** be asked.

Remember to ask the participant to quantify items, such as frequency, if they do not spontaneously do so.

1. *Hesitancy - also considering any differences in occurrence with an empty, half full or over full bladder.*

When you go to urinate, do you find that you have difficulty with starting to urinate?
- Is this different depending on how full you feel that your bladder is?
- Must clarify if it is harder or easier, as the bladder is fuller.

2. *Straining - requirement to start flow, to maintain flow and whether they can use it to increase flow.*

When you go to urinate, do you feel you have to strain or squeeze, or push the urine out?
- Could you describe to me how you strain or squeeze?
- Do you do this to start urinating, or to maintain your stream, or both?
- Females: Do you have to push your uterus up to be able to urinate?

Why do you attempt to strain like this?
- Need to clarify if they strain to prevent an unwanted outcome, or if they feel they still have urine left to void.

3. *Incomplete emptying after finish passing urine - does it feel like they've emptied their bladder.*

After you've finished urinating, do you feel that you have emptied your bladder completely?
- How soon do you feel the need to urinate again afterwards?

4. *Flow - whether continuous and if so any fluctuations between weak and strong during flow, if interrupted - does it stop and start during the void.*

How would you describe your urine flow?
- Is the urine flow continuous, or does it stop and start?
  - How long do you wait for it to start again?
  - How many times does it stop and start during a urination?
- If the flow is continuous, does the strength of the stream go up and down?
- Has it always been this way, before you developed your condition?

5. *Bladder sensation - do these individuals reach a strong desire to void if they hold their urine. It is normal to feel a strong desire to void if urine is held for too long but some individuals do not experience this even though they are very/over full and we need to know what experience of sensations DU/UAB patients report.*

If you hold your urine in, do you feel a strong desire to urinate?
- How soon does this happen, after you first feel the need to go?

6. *Cluster of urinations - passing urine at regular intervals or at different times of the day is their frequency increased, and if so, when.*

Are there times of the day when you pass urine more frequently?
Are you aware of having to wait after you think you have finished urinating?

This could be while remaining seated, or dressing, or leaving the toilet and returning. Ask the participant to clarify.

Differentiate between: the distribution during the day, and clustering per episode (for example, toilet hanging).

7. General voiding questions – some participants have described a variability in their voiding experiences. These questions aim to address that.

Could you describe what your most difficult urination is like?

Could you describewhat your easiest urination is like?

Of these two experiences, which occurs the most regularly?

8. Leakage – to check that the participants do not leak, but if they do then how much.

Do you ever leak urine?

If so, in what context?

How much leakage occurs?

How often?

Exploring IMPACT (10 MINUTES)

Exploration of specific areas of impact

Thinking back to before you developed these symptoms, has your life changed since?

- If they try to do things normally, explore their adaptive strategies.

How does your bladder condition affect your relationships with the other people in your life?

How does it affect the types of things that you do?

- Are there areas of your life that you feel excluded from?

How does it affect the way you feel about yourself?

- Embarrassment, concentration, motivation, satisfaction/happiness, stress/anxiety.

How does it affect your ability to sleep, and get rest?

- Recovery, daytime function.

Since developing your condition, are you more aware of looking for the toilet when you are out of the house?

Importance of areas of impact (5 MINUTES)

"Thinking about the issues we have been talking about today, which are the most important to you?"

If the participant is having difficulty with recall (either too many or too few points), then ask participants to select the **three** most important issues for them

Summing up (5 MINUTES)

"As this is an opportunity for you to have your individual voice heard, is there anything we haven't covered that you would like to discuss?"

Remind participant that everything he/she has said will be kept confidential, and also the need for confidentiality from them.

Close the interview

> "Thank you for taking the time to share your thoughts and experiences with us. If you have any further concerns or questions following on from today, please do not hesitate to contact me. Thank you again."

Appendix 4 Patient consent form

ICIQ Study

Consent Form

Centre Number:

Study Number:

Patient Information Number for this trial:

CONSENT FORM

Title of Project: Development of the ICIQ Questionnaire

Name of Researcher: Alan Uren

Please initial box ☐

1. I confirm that I have received enough information about this study. ☐

2. I understand that my participation is voluntary and that I am free to withdraw at any time without my medical care or legal rights being affected.

3. I am willing to allow access to relevant sections of my medical notes and understand that data collected during the study may be looked at by individuals from ICIQ, from regulatory authorities or from the NHS Trust, where it is relevant to my taking part in this research but understand that strict confidentiality will be maintained. The purpose of this is to check that the study is being carried out correctly. ☐

4. I agree to take part in the above study. ☐

_____	_____	_____
Name of patient	Date	Signature
_____	_____	_____
Name of person taking consent (if different from researcher)	Date	Signature
_____	_____	_____
Researcher	Date	Signature

Appendix 5 ICIQ-UAB developmental version 1

Initial number

CONFIDENTIAL

DAY MONTH YEAR
Today's date

Underactive Bladder

Many people experience urinary symptoms some of the time. This questionnaire aims to find out whether you experience symptoms known to be associated with underactive bladder, and whether or not these symptoms have an impact on your everyday life. We would be grateful if you could answer the following questions, thinking about how your symptoms have been, on average, over the **LAST 24 HOURS**.

1 Please write in your date of birth:

DAY MONTH YEAR

2 Are you *(tick one)***:** Female ☐ Male ☐

3a. How often do you pass urine during the day, on average?

1-6 times ☐	0
7-8 times ☐	1
9-10 times ☐	2
11-12 times ☐	3
13 or more times ☐	4

3b. How much does this bother you?
Please ring a number between 0 (not at all) and 10 (a great deal)

0 1 2 3 4 5 6 7 8 9 **10**
not at all — a great deal

4a. Is there a delay before you can start to urinate?

never ☐	0
occasionally ☐	1
sometimes ☐	2
most of the time ☐	3
all of the time ☐	4

4b. How much does this bother you?
Please ring a number between 0 (not at all) and 10 (a great deal)

0 1 2 3 4 5 6 7 8 9 **10**
not at all — a great deal

5a. How often do you feel you have to concentrate in order to start urinating?

never	0
occasionally	1
sometimes	2
most of the time	3
all of the time	4

5b. How much does this bother you?
Please ring a number between 0 (not at all) and 10 (a great deal)

0 1 2 3 4 5 6 7 8 9 **10**
not at all a great deal

6a. Do you ever leak urine before you can get to the toilet?

never	0
occasionally	1
sometimes	2
most of the time	3
all of the time	4

6b. How much does this bother you?
Please ring a number between 0 (not at all) and 10 (a great deal)

0 1 2 3 4 5 6 7 8 9 **10**
not at all a great deal

7a. During the night, how many times do you have to get up to urinate, on average?

none	0
one	1
two	2
three	3
four or more	4

7b. How much does this bother you?
Please ring a number between 0 (not at all) and 10 (a great deal)

0 1 2 3 4 5 6 7 8 9 **10**
not at all a great deal

8a. How often do you feel you bladder has not emptied properly after you have urinated?

never ☐ 0
occasionally ☐ 1
sometimes ☐ 2
most of the time ☐ 3
all of the time ☐ 4

8b. How much does this bother you?
Please ring a number between 0 (not at all) and 10 (a great deal)

0 1 2 3 4 5 6 7 8 9 **10**
not at all a great deal

9a. After you have finished urinating how often do you feel like you could go again even though you can't?

never ☐ 0
occasionally ☐ 1
sometimes ☐ 2
most of the time ☐ 3
all of the time ☐ 4

9b. How much does this bother you?
Please ring a number between 0 (not at all) and 10 (a great deal)

0 1 2 3 4 5 6 7 8 9 **10**
not at all a great deal

10a. Have you ever blocked up completely so that you could not urinate at all and had to have a catheter passed to drain the bladder?

no ☐ 0
yes, once ☐ 1
yes, twice ☐ 2
Yes, more than twice ☐ 3

10b. How much does this bother you?
Please ring a number between 0 (not at all) and 10 (a great deal)

0 1 2 3 4 5 6 7 8 9 **10**
not at all a great deal

11a. How often do you need to self-catheterise?

never ☐ 0
occasionally ☐ 1
sometimes ☐ 2
most of the time ☐ 3
all of the time ☐ 4

11b. How much does this bother you?
Please ring a number between 0 (not at all) and 10 (a great deal)

0 1 2 3 4 5 6 7 8 9 **10**
not at all — a great deal

12a. Do you have to strain or squeeze to urinate?

never ☐ 0
occasionally ☐ 1
sometimes ☐ 2
most of the time ☐ 3
all of the time ☐ 4

12b. How much does this bother you?
Please ring a number between 0 (not at all) and 10 (a great deal)

0 1 2 3 4 5 6 7 8 9 **10**
not at all — a great deal

13a. Do you have to strain to <u>start</u> urinating?

never ☐ 0
occasionally ☐ 1
sometimes ☐ 2
most of the time ☐ 3
all of the time ☐ 4

13b. How much does this bother you?
Please ring a number between 0 (not at all) and 10 (a great deal)

0 1 2 3 4 5 6 7 8 9 **10**
not at all — a great deal

14a. Do you have to strain to try and improve the flow of your urination?

never ☐ 0
occasionally ☐ 1
sometimes ☐ 2
most of the time ☐ 3
all of the time ☐ 4

14b. How much does this bother you?
Please ring a number between 0 (not at all) and 10 (a great deal)

0 1 2 3 4 5 6 7 8 9 **10**
not at all a great deal

15a. How often do you strain or squeeze at the <u>end</u> of your urination?

never ☐ 0
occasionally ☐ 1
sometimes ☐ 2
most of the time ☐ 3
all of the time ☐ 4

15b. How much does this bother you?
Please ring a number between 0 (not at all) and 10 (a great deal)

0 1 2 3 4 5 6 7 8 9 **10**
not at all a great deal

16a. How often do you experience a sudden urge or desire to urinate which you are <u>unable</u> to ignore?

never ☐ 0
occasionally ☐ 1
sometimes ☐ 2
most of the time ☐ 3
all of the time ☐ 4

16b. How much does this bother you?
Please ring a number between 0 (not at all) and 10 (a great deal)

0 1 2 3 4 5 6 7 8 9 **10**
not at all a great deal

17a. Have you had a urinary tract infection in the past month?

no ☐ 0

yes ☐ 1

17b. How much does this bother you?

Please ring a number between 0 (not at all) and 10 (a great deal)

0 1 2 3 4 5 6 7 8 9 **10**

not at all — a great deal

18a. Would you say that the strength of your urinary stream is…

normal ☐ 0

occasionally reduced ☐ 1

sometimes reduced ☐ 2

reduced most of the time ☐ 3

reduced all of the time ☐ 4

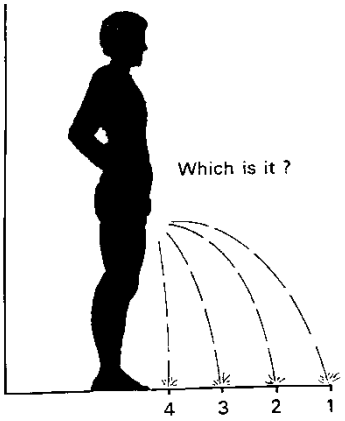
18b. How much does this bother you?

Please ring a number between 0 (not at all) and 10 (a great deal)

0 1 2 3 4 5 6 7 8 9 **10**

not at all — a great deal

19. Would you say that the strength of your urinary stream is...
(please ring one number)



(from Peeling, 1989)

20a. Do you stop and start more than once while you urinate?

never ☐ 0
occasionally ☐ 1
sometimes ☐ 2
most of the time ☐ 3
all of the time ☐ 4

20b. How much does this bother you?
Please ring a number between 0 (not at all) and 10 (a great deal)

0 1 2 3 4 5 6 7 8 9 **10**
not at all a great deal

21a. How often do a few drops leak out into your underwear shortly after you have finished urinating and have dressed yourself?

never ☐ 0
occasionally ☐ 1
sometimes ☐ 2
most of the time ☐ 3
all of the time ☐ 4

21b. How much does this bother you?
Please ring a number between 0 (not at all) and 10 (a great deal)

0 1 2 3 4 5 6 7 8 9 **10**
not at all a great deal

22a. Do you experience pain in your bladder?

never ☐ 0
occasionally ☐ 1
sometimes ☐ 2
most of the time ☐ 3
all of the time ☐ 4

22b. How much does this bother you?
Please ring a number between 0 (not at all) and 10 (a great deal)

0 1 2 3 4 5 6 7 8 9 **10**
not at all a great deal

23a. If you have to wait to go to the toilet do you have pain in your bladder?

never ☐ 0
occasionally ☐ 1
sometimes ☐ 2
most of the time ☐ 3
all of the time ☐ 4

23b. How much does this bother you?
Please ring a number between 0 (not at all) and 10 (a great deal)

0 1 2 3 4 5 6 7 8 9 **10**
not at all a great deal

24a. Do you have to urinate again (within 15 minutes) after you thought you had finished urinating?

never ☐ 0
occasionally ☐ 1
sometimes ☐ 2
most of the time ☐ 3
all of the time ☐ 4

24b. How much does this bother you?
Please ring a number between 0 (not at all) and 10 (a great deal)

0 1 2 3 4 5 6 7 8 9 **10**
not at all a great deal

25a. How often do you feel that you have no control over the direction of your stream?

never ☐ 0
occasionally ☐ 1
sometimes ☐ 2
most of the time ☐ 3
all of the time ☐ 4

25b. How much does this bother you?
Please ring a number between 0 (not at all) and 10 (a great deal)

0 1 2 3 4 5 6 7 8 9 **10**
not at all a great deal

26a. **How often do you feel that you were not able to pass what you might consider a satisfactory amount of urine?**

never	☐	0
occasionally	☐	1
sometimes	☐	2
most of the time	☐	3
all of the time	☐	4

26b. **How much does this bother you?**
Please ring a number between 0 (not at all) and 10 (a great deal)

0 1 2 3 4 5 6 7 8 9 **10**

not at all — a great deal

The following questions relate to how these symptoms may have affected your everyday life. Please answer each question according to how often you have felt this way over the **PAST 4 WEEKS.**

27a. How much do you have to plan your life around the location of toilets?

never	☐	0
occasionally	☐	1
sometimes	☐	2
most of the time	☐	3
all of the time	☐	4

27b. How much does this bother you?
Please ring a number between 0 (not at all) and 10 (a great deal)

0 1 2 3 4 5 6 7 8 9 **10**
not at all — a great deal

28a. How often do you feel that your urinary problem interferes with your social life?

never	☐	0
occasionally	☐	1
sometimes	☐	2
most of the time	☐	3
all of the time	☐	4

28b. How much does this bother you?
Please ring a number between 0 (not at all) and 10 (a great deal)

0 1 2 3 4 5 6 7 8 9 **10**
not at all — a great deal

29a. How often do your symptoms prevent you from getting the amount of sleep you needed?

never	☐	0
occasionally	☐	1
sometimes	☐	2
most of the time	☐	3
all of the time	☐	4

29b. How much does this bother you?
Please ring a number between 0 (not at all) and 10 (a great deal)

0 1 2 3 4 5 6 7 8 9 **10**
not at all — a great deal

30a. **How often do you feel tired the next day because of having to get up at night to urinate?**

never ☐	0
occasionally ☐	1
sometimes ☐	2
most of the time ☐	3
all of the time ☐	4

30b. **How much does this bother you?**
Please ring a number between 0 (not at all) and 10 (a great deal)

0 1 2 3 4 5 6 7 8 9 **10**
not at all a great deal

31. **Overall, how much impact does getting up at night to urinate have on your day to day life?**

Please ring a number between 0 (not at all) and 10 (a great deal)

0 1 2 3 4 5 6 7 8 9 **10**
not at all a great deal

32a. **How often do you feel that your urinary problem limits your physical activities (e.g. exercise, sport)?**

never ☐	0
occasionally ☐	1
sometimes ☐	2
most of the time ☐	3
all of the time ☐	4

32b. **How much does this bother you?**
Please ring a number between 0 (not at all) and 10 (a great deal)

0 1 2 3 4 5 6 7 8 9 **10**
not at all a great deal

33a. Does your urinary problem affect the way you feel about yourself?

never	☐	0
occasionally	☐	1
sometimes	☐	2
most of the time	☐	3
all of the time	☐	4

33b. How much does this bother you?
Please ring a number between 0 (not at all) and 10 (a great deal)

0 1 2 3 4 5 6 7 8 9 **10**
not at all a great deal

34a. Does your urinary problem cause you to feel embarrassed?

never	☐	0
occasionally	☐	1
sometimes	☐	2
most of the time	☐	3
all of the time	☐	4

34b. How much does this bother you?
Please ring a number between 0 (not at all) and 10 (a great deal)

0 1 2 3 4 5 6 7 8 9 **10**
not at all a great deal

35a. Are you careful how much fluid you drink?

never	☐	0
occasionally	☐	1
sometimes	☐	2
most of the time	☐	3
all of the time	☐	4

35b. How much does this bother you?
Please ring a number between 0 (not at all) and 10 (a great deal)

0 1 2 3 4 5 6 7 8 9 **10**
not at all a great deal

36a. **Are your urinary problems affected by your bowel movements?**

never	☐	0
occasionally	☐	1
sometimes	☐	2
most of the time	☐	3
all of the time	☐	4

36b. **How much does this bother you?**
Please ring a number between 0 (not at all) and 10 (a great deal)

0 1 2 3 4 5 6 7 8 9 **10**
not at all a great deal

37. **Overall, how much would you say your urinary symptoms interfere with your everyday life?**

Please ring a number between 0 (not at all) and 10 (a great deal)

0 1 2 3 4 5 6 7 8 9 **10**
not at all a great deal

Thank you very much for answering these questions.

Appendix 6 Cognitive interview schedule

UAB PRO CD interview schedule v3 (27/02/15)

Firstly I will ask you to fill in the questionnaire. This will be timed in order for us to have an idea of how long it takes to complete. After this I will ask you a few general questions about the questionnaire. We will then go through each question in more detail asking for your comments on each individual item. Please feel free to be as open and honest as you like when you tell me your thoughts and comments. This is a brand new questionnaire so any input you have is extremely valuable to us. It will enable us to improve the questionnaire and make it clear, relevant and easy to complete in the future.

To be completed shortly after initial completion of the questionnaire by the participant.

Completion instruction Page 1	**Purpose**: To understand respondent's interpretation of the task (s) to be performed.			
	Intended interpretation	Participant's interpretation	Comments	Suggestions for change
Can you tell me in your own words, what this instruction is asking you to do?				
Can you describe any confusion or difficulty you had in understanding these instructions?				
Are there any words or phrases that you would change to improve the instructions.				
Recall timeframe Page 1	**Purpose: To identify how patients retrieve information, remember situations or events.**			
	Intended interpretation	Participant's interpretation	Comments	Suggestions for change
What does (timeframe e.g 'past 24 hours') mean to you?	Recall symptoms over the last day			
What period of time did you think about when you were completing the questionnaire?				
Describe your experiences with your own symptoms over this period (timeframe).				

Completion instruction Page 10	**Purpose**: To understand respondent's interpretation of the task (s) to be performed.			
	Intended interpretation	Participant's interpretation	Comments	Suggestions for change
Can you tell me in your own words, what this instruction is asking you to do?				
Can you describe any confusion or difficulty you had in understanding these instructions?				
Are there any words or phrases that you would change to improve the instructions.				
Recall timeframe Page 10	**Purpose: To identify how patients retrieve information, remember situations or events.**			
	Intended interpretation	Participant's interpretation	Comments	Suggestions for change
What does (timeframe e.g 'past 24 hours') mean to you?	Recall symptoms over the last month			
What period of time did you think about when you were completing the questionnaire?				
Describe your experiences with your own symptoms over this period (timeframe).				

Length	**Purpose**: To determine if the length of time it takes to complete the questionnaire is reasonable (does not burden subject).	
	Comments	Suggestions for change
What did you think about the amount of time it took you to complete the questionnaire?		
If the questionnaire were reduced to three quarters/half the length would this be a burden?		
Content coverage	**Purpose**: To determine if the content in the instrument is comprehensive/to assure that there are no missing concepts.	
	Comments	Suggestions for change
What other symptoms do you have with DU/UAB that are not covered in this questionnaire?		
What other experiences do you have with DU/UAB that are not covered in this questionnaire?		

Format	**Purpose**: To identify respondent difficulties with the presentation of the questionnaire or diary.	
	Comments	Suggestions for change
What suggestions do you have for changing the questionnaire so it is easier to complete?		
Observation of questionnaire completion: note facial expressions, indications of reading difficulty, flipping pages back and forth. Listen for comments about difficulty reading or questions that indicate lack of clarity or ease of use.		

Item stem Q3a	**Purpose**: To understand the clarity of the question from the respondent's perspective.			
	Intended interpretation	Participant's interpretation	Comments	Suggestions for change
Using your own words, how would you explain what this question means?	The complaint by the patient that he or she urinates too often by day			
Could you interpret this question in a different way?				
Please explain any difficulties you experience with this question?				
Response frame Q3a	**Purpose**: To understand how participants interpret the response options and make decisions around response choice.			
	Intended interpretation	Participant's interpretation	Comments	Suggestions for change
Please read each response choice and tell me what it means to you.				
In thinking about your experience with [specific concept of interest],which response best describes your experience?				
What caused you to choose this response? Would you ever choose [first response in scale]? Why or why not? Can you describe an experience where you might choose [last response in scale]?				
Bother question Q3b Is this relevant and appropriate to this question?				

Appendix 7 Revisions to the draft ICIQ-UAB following results of the pilot study

Introduction

A pilot study was conducted to test the instrument measurement properties, detailed in chapter 6. Subsequently, minor changes were implemented to be tested on the basis of this new psychometric evidence. The following report details the rationale for the changes made to versions 11b of the ICIQ-UAB, used in the pilot study. The resulting version 12 was tested with patients in seven additional cognitive) interviews conducted over the phone. This investigated the facility of understanding and interpretation of these changes in the target population. Decisions were made surrounding their retention in version 13.

Revisions and rationale for changes to the ICIQ-UAB v11b and ICIQ-UAB v12

The following summarises the rationale for changes made to ICIQ-UAB version 11b and version 12. A summary of the patient input during the additional cognitive interviews is included when relevant. Floor effect percentages cited are from the pilot study at baseline (percentage of responses with lowest response option at day 1 in the 1 week recall period version).

Additional cognitive interview sample

All participants for the additional cognitive interviews were recruited from the pool of participants who had previously participated in the original qualitative phase CE and/or cognitive interviews. Five male and two female participants, with or without a range of coexisting conditions, were purposively selected in order to ensure multiple gender and participant perspectives on the changes made (Table 35). Informed consent was taken over the phone and all were audio recorded. Participants were emailed the draft questionnaire in advance of the call. Notes were taken during the interviews using the interview schedule included in the Appendices.

Table 35. Participant characteristics for the additional cognitive interviews.

Additional CD interview number	CD and/or CE study number	Group	Gender	Age (years)	PVR	ISC
01	CE P8 CD P21	DU + USI	F	70	Yes	No
02	CD P6	DU+ BOO + DO	M	81	Yes	No
03	CD P4	DU + DO + BOO	M	67	Yes	No
04	CE P23 CD P30	DU	F	75	Yes	Yes, twice a day
05	CD P5	DU	M	69	Yes	Yes, every day
06	CD P3	DU + DO +BOO-E	M	69	Yes	No
07	CD P8	DU + DO + SUI	M	75	Yes	Yes, occasionally

Abbreviations: Detrusor underactivity (DU), detrusor overactivity (DO), stress urinary incontinence (SUI), bladder outlet obstruction (BOO), bladder outlet obstruction in the equivocal range (BOO-E), Post void residual (PVR) >30ml, Intermittent self-catheterisation (ISC).

Questionnaire content and aspects of completion

General changes

Clinical experts commenting on the draft version of the instrument felt the term ‘urinate’ was generally avoided and ‘pass urine’ was preferred, based on many years of clinical experience. Given that participants

involved in the additional cognitive interviews had no preference, it was decided to replace this term throughout the instrument.

The term 'urinary symptoms' was also considered to be colloquially preferred due to the negative connotations associated with 'urinary problems' so this was replaced in all items in which it occurred.

Recall period

Both versions were very similar in terms of the psychometric properties which they exhibited. Three more items in the ICIQ-UAB 24hrs had floor effects that were marginally over the threshold of 20%. However, the highest response option was not used in three more of the items in ICIQ-UAB (1 week version). The internal consistency and test-retest identical agreement between repeated administrations was very similar for both versions but there was correlational evidence that some items could be more suitable for a shorter recall periods. There was no strong evidence to suggest an overall advantage of either recall period in the pilot study. However, the original and additional cognitive interviews suggested patients preferred a longer recall period for the impact items to reflect the length of time that patients required to notice a discernible impact. A shorter recall period is recommended by the Food and Drug Agency guidance for industry as it reduces recall bias [99] so the 24 hour recall period was preferred for the symptoms and sign items for future clinical trials. A longer recall period of 1 week was retained for the impact items.

Response options

Discussions among the development team suggested that the middle response option of 'sometimes' may lack specificity and it was therefore suggested to evaluate the alternative response option of 'about half the time', as incorporated in the 24 hour version. Patients in the additional cognitive interviews were asked if they had any preference between 'sometimes' and 'about half the time' for the third response option. The option 'about half the time' was perceived as being more precise, but both options were understood as intended. The decision was taken to retain 'about half the time' due to the compatibility with the 24 hour recall period used in version 13.

Version 11b	Version 12	Version 13
Not at all occasionally sometimes most of the time every time	Not at all occasionally about half the time most of the time all of the time	Retained

Bother score items

The pilot study demonstrated a clear relationship between high bother scores and reported severity of condition. The bother scores were therefore an effective measure of the 'bother' relating to each symptom, sign or impact. No changes were made to the associated bother questions for each item.

Medical history items

Acute retention

This item is not part of the scoring as will not change over the study period, but despite a strong floor effect (83%) it was considered clinically relevant as a sign and for health care resource utilization. This question is to be considered for removal from the tool at a later stage, with the possible inclusion in a baseline questionnaire or case report form (CRF) as part of a study, or separately in clinical use.

Version 11b	Version 12	Version 13
Item 1	Item 1	Item 1
Have you ever been unable to urinate at all and had to go to hospital to have a catheter tube inserted to drain your urine?	Retained	Retained
No Once Twice Three or more times	Retained	Retained

Urinary tract infection

This item is not part of the scoring as will not change over the study period due to the incorporated recall period. Despite a strong floor effect (60%) it collects relevant clinical information at baseline and is included in the questionnaire at this stage with the same proposed approach as question 1.

Version 11b	Version 12	Version 13
Item 2	Item 2	Item 2
Over the last 12 months, have you had a urinary infection for which you took medication?	Retained	Retained
No Unsure Once Twice Three or more times	Retained	Retained

Self-catheterisation

The pilot study day 1 data returned a strong floor effect for this item (80%).The recall period for this item was changed to a shorter period of 1 week in order to capture the information supplied by the target population and to be considered more clinically relevant. The associated response options were changed to a numerical representation and to be compatible with the new recall period. These were confirmed to be well understood and acceptable by patients in the additional cognitive interviews.

Version 11b	Version 12	Version 13
Item 3	Item 3	Item 3
Over the last month, how often did you self-catheterise?	Over the last week, how often did you self-catheterise?	Retained
Not at all about once a week or less often two or three times a week about once a day more than once a day every time you urinated	Not at all less than once a day (1-6 times per week) 1-2 times a day 3-4 times a day 5 or more times a day	Retained

Symptom items

Hesitancy

There was some concern due to the evidence of clustering among the middle response options that the current wording of the question was capturing non-symptomatic patients so a proposed change to the wording was tested in version 12. However, the patients in the additional cognitive interviews preferred the brevity and wording of version 11b as 'more than a few seconds' was considered to be 'too prescriptive'. The patient also confirmed the 'delay' to only be reported when symptomatic. Version 11b was retained.

Version 11b	Version 12	Version 13
Item 4	Item 4	Item 4
When ready to urinate, was there a delay before the urine flow started?	When ready to pass urine, did you have to wait more than a few seconds for the urine flow to start?	When ready to pass urine, was there a delay before the urine flow started?
Not at all occasionally sometimes most of the time every time	Not at all occasionally about half the time most of the time every time	Retained

Concentrate to start void

This item had a high correlation with item 4 and item 12 (Pearson correlation coefficient 0.79 and 0.75 respectively) indicating potential redundancy within these items. However, further data from responsiveness to change data will add to evidence for inclusion or removal. This question item was retained with no change.

Version 11b	Version 12	Version 13
Item 5	Item 5	Item 5
When ready to urinate, did you feel you had to concentrate in order to start urinating?	Retained	Retained
Not at all occasionally sometimes most of the time every time	Not at all occasionally about half the time most of the time every time	Retained

Small volume of urine per void

This item showed a relatively even distribution of answers over the response options in the pilot study. The decision was made to keep the item unchanged.

Version 11b	Version 12	Version 13
Item 6	Item 6	Item 6
How often were you only able to pass a small volume of urine?	Retained	Retained
Not at all occasionally sometimes most of the time every time	Not at all occasionally about half the time most of the time every time	Retained

Post micturition dribble

A high floor effect for this item was found in the pilot study data (50%). The item was left unchanged awaiting further responsiveness to change data.

Version 11b	Version 12	Version 13
Item 7	Item 7	Item 7
How often did a few drops leak out into your underwear shortly after you had finished urinating and had dressed	Retained	Retained

yourself?		
Not at all occasionally sometimes most of the time every time	Not at all occasionally about half the time most of the time every time	Retained

Incontinence

A very high floor effect (70%) was found in the pilot study for this item and the highest response option of 'every time' was not utilized. In the CE interviews incontinence was more associated with DU and coexisting conditions so there may be a low DU specificity. However as this symptom is very bothersome to those in which it occurs and to leak 'every time' is theoretically possible, the item is retained unchanged, but may be potentially removed following evidence from the responsiveness to change phase of the study.

Version 11b	Version 12	Version 13
Item 8	Item 8	Item 8
How often did you leak urine (e.g. before you could get to the toilet, when physically active, or when you coughed or sneezed)?	Retained	Retained
Not at all occasionally sometimes most of the time every time	Not at all occasionally about half the time most of the time every time	Retained

Need to immediately re-void

A small floor effect was found in the pilot study data (28%). However, as the distribution of responses was otherwise satisfactory, the decision was made to retain this item unchanged.

Version 11b	Version 12	Version 13
Item 9	Item 9	Item 9
After you had urinated, how often did you have to return to the bathroom to urinate again, within a short space of time (e.g. within 15 minutes)?	Retained	After passing urine, how often did you have to return to the bathroom to pass urine again, within a short space of time (e.g. within 15 minutes)?
Not at all occasionally sometimes most of the time every time	Not at all occasionally about half the time most of the time every time	Retained

Sensation of incomplete emptying

No floor effect in this item (<20%) was observed and there was a reasonable division over the response options, suggesting a good potential for responsiveness to change. No changes were made to this item.

Version 11b	Version 12	Version 13
Item 10	Item 10	Item 10
Soon after you had urinated, how often did you have a sensation that your bladder was not completely empty?	Retained	Retained
Not at all occasionally sometimes most of the time every time	Not at all occasionally about half the time most of the time every time	Retained

Intermittency

There was discussion by the development team surrounding whether 'start and stop' made more intuitive sense than 'stop and start'. Patients had a preference for the perceived straightforwardness of version 12. This was retained with the addition of 'how often' for consistency.

Version 11b	Version 12	Version 13
Item 11	Item 11	Item 11
How often did you stop and start more than once, during your urinations?	When you passed urine, did the flow stop and start?	When you passed urine, how often did the flow stop and start?
Not at all occasionally sometimes most of the time every time	Not at all occasionally about half the time most of the time every time	Retained

Straining

As a result of expert clinical opinion, an additional straining item was included in version 12 to ascertain if the patient strains to maintain the flow while passing urine. Following additional cognitive interview testing, each straining question (to start, maintain, or finish) was considered mutually exclusive by most of the interviewed patients and were answered as separate items. The new item was interpreted as expected by all patients in the additional cognitive interviews. All three questions relating to straining were included in version 13.

Version 11b	Version 12	Version 13
Item 12	Item 12	Item 12
How often did you strain to start your urinations?	How often did you strain to start passing urine?	Retained
Not at all occasionally sometimes most of the time every time	Not at all occasionally about half the time most of the time every time	Retained

Version 11b	Version 12	Version 13
No corresponding item	Item 13	Item 13
	When passing urine, how often did you strain to maintain your flow?	Retained
	Not at all occasionally about half the time most of the time every time	Retained

Version 11b	Version 12	Version 13
Item 13	Item 14	Item 14
How often did you strain towards the end of your urinations to try and empty your bladder?	How often did you strain towards the end of passing urine, to try and empty your bladder?	Retained
Not at all occasionally sometimes most of the time every time	Not at all occasionally about half the time most of the time every time	Retained

Slow stream

There was an even distribution of responses for each option for this item apart from the highest category. No changes were made to this item.

Version 11b	Version 12	Version 13
Item 14	Item 15	Item 15
On average, would you say that the strength of flow of your urinary stream was…	Retained	Retained
Normal (not reduced) A little reduced Reduced Very reduced Extremely reduced	Retained	Retained

Urgency

There was a floor effect (50% in groups 0 or 1) for this item. There was some discussion that the highest response option 'every time' should be rephrased or is not required due to this being very unlikely to occur in this patient population. However, the decision was made to keep the item unchanged as theoretically this is possible, and therefore an option for 'every time' should be included.

Version 11b	Version 12	Version 13
Item 15	Item 16	Item 16
How often did you experience a sudden or strong need to pass urine which you were unable to ignore, and had to rush to the bathroom?	Retained	Retained
Not at all occasionally sometimes most of the time every time	Not at all occasionally about half the time most of the time every time	Retained

Nocturia/nocturnal voids

Following discussion with the development team, version 11b was retained because it was deemed clinically more useful to measure nocturnal voids (rather than actual nocturia i.e 'to awake and get up'). This symptom is known to have complex aetiologies [47] so there was some doubt that it would show responsiveness to change for a DU specific intervention. However, it is considered important for inclusion as is highly bothersome to many patients and often mentioned spontaneously in the CE study.

Version 11b (1 week version)	Version 12	Version 13
Item 16	Item 17	Item 17
During the night, how many times did you have to get up to urinate, on average?	During the night, how many times did you have to get up to pass urine?	Retained
Not at all Once Twice Three times Four or more times	Retained	Retained

Daytime urinary frequency

The scoring of this item was under scrutiny by the development team due to discussion around what may be considered a 'normal' urinary frequency. Although context dependent, it was considered that '4-8 times' would provide a sufficiently broad response option to capture natural variation for non-symptomatic individuals. This was implemented in version 13 following testing with patients in the additional cognitive interviews. The response option '1-3 times' is a relatively low daytime urinary frequency so may be considered symptomatic. For this reason, treatment may cause patients to improve in either direction. This is reflected in the alternative proposed scoring of these response options, although this will require full statistical evaluation to establish its appropriateness.

Version 11b	Version 12	Version 13	Scoring
Item 17	Item 18	Item 18	
During the day, how many times did you urinate, on average?	During the day, how many times did you pass urine, on average?	During the day, how many times did you pass urine?	
1-3 times 4-6 times 7-9 times 10-12 times 13 or more times	1-3 times 4-7 times 8-9 times 10-11 times 12 or more times	1-3 times 4-8 times 9-10 times 11-12 times 13 or more times	1 0 1 2 3

Reduced sensation of the fullness of the bladder

A slight change in wording was proposed by the development team in order to aid clarity of interpretation. A comparison of the difference in wording between version 11b and version 12 by patients in the additional cognitive interviews appeared to result in very little or no difference in interpretation. However, there was

a slight preference for the wording of version 12 by patients so this was retained. In addition, the response options were changed to measure frequency following confirmation that these were interpreted as intended in the additional cognitive interviews.

Version 11b	Version 12	Version 13
Item 18	Item 19	Item 19
Did you find it difficult to tell how full your bladder was?	Did you find it difficult to tell when your bladder was full?	Retained
Not difficult a little difficult difficult very difficult extremely difficult	Not at all occasionally about half the time most of the time all of the time	Retained

Waiting in bathroom after voiding

Following discussion at the development team meeting slight changes to the wording were suggested to aid clarity of the item. Due to patient preference in the additional cognitive interviews, the wording was changed from 'wait' to 'stay', and 'a bit longer' was retained in version 13.

Version 11b	Version 12	Version 13
Item 19	Item 20	Item 20
How often did you wait a bit longer in the bathroom after urinating, to make sure your bladder was as empty as possible?	How often did you wait in the bathroom after passing urine, to make sure your bladder was as empty as possible?	How often did you stay a bit longer in the bathroom after passing urine, to make sure your bladder was as empty as possible?
Not at all occasionally sometimes most of the time every time	Not at all occasionally about half the time most of the time all of the time	Retained

Length of time in bathroom

A revision to the wording to aid clarity was suggested following discussion at the development team meeting. This was well received by the patients in the cognitive interviews. The change in response options were also thought to better reflect the time in which patients tended to spend in the bathroom, and would allow more honest answers.

Version 11b	Version 12	Version 13
Item 20	Item 21	Item 21
What was the longest time that you needed to spend in the bathroom to finish urinating?	What was the longest time that you needed to spend in the bathroom trying to empty your bladder?	Retained
less than a minute 1-5 minutes 6-10 minutes 11-15 minutes more than 15 minutes	1-2 minutes 3-5 minutes 6-10 minutes 11-15 minutes more than 15 minutes	Retained

Temporarily unable to pass urine

There was a large floor effect observed in the pilot study for this item (61%) and the highest two scores were not recorded by any of the patients. All patients in the additional cognitive interviews confirmed this symptom as infrequent, if it occurred at all, so were likely only to answer 'occasionally', even if the recall period was 1 week. However, the average bother score for those who had these symptoms in the pilot study was relatively high (6). It was considered to have potentially important clinical relevance and no further suggestions were made for the wording or response options. This item may be considered for removal or further amendment following the responsiveness to change data.

Version 11b	Version 12	Version 13
Item 21	Item 22	Item 22
How often did you go to the toilet to pass urine but were unable to urinate at all, so had to return to the bathroom to try again later?	Retained	Retained
Not at all occasionally sometimes most of the time every time	Not at all occasionally about half the time most of the time every time	Retained

Bowel associations

Seventy percent of participants reported no problems with their bowels, and for the patients who have problems, the severity was low. In addition, associations with bowel problems were not deemed of sufficient clinical relevance. This item was removed for version 12.

Version 11b	Version 12	Version 13
Item 22	Removed	
Did you have problems with your bowels? (yes/no) If so, were your urinary symptoms made worse by this?		
Not at all occasionally sometimes most of the time every time		

Clustering of symptoms

No subjects ticked multiple boxes here and this was the only item with missing data >5%. In addition, it was deemed to have limited clinical importance. This item was removed for version 12.

Version 11b	Version 12	Version 13
Item 23	Removed	
Were your urinary symptoms worse at particular times? Please tick those that apply…		
No In the morning In the afternoon In the evening At night		

Impact items

The impact items were all observed to have a high floor effect in the pilot study (30-50%). In order to improve the capacity for responsiveness to change a number of changes were made to the impact items as detailed below.

Planning life around toilet visits

Following discussion at the development team meeting it was felt that examples should be given to aid clarity on the intention of the question. These were confirmed by the cognitive interview patients to be helpful so were retained in version 13. 'Plan your activities' was also tested versus 'make plans' and the latter was thought to allow greater flexibility in answering, so was retained.

Version 11b	Version 12	Version 13
Item 24	Item 23	Item 23
How often did you plan your life around the location of toilets?	How often did you make plans around the location of toilets (e.g. shopping, social outings, travelling, holidays)?	Retained
Not at all occasionally sometimes most of the time every time	Retained	Retained

Social life

This question was reworded to include social life and work as part of 'day to day life' following discussion with the development team. There was some concern from patients in the additional cognitive interviews that this question overlapped with the 'planning life around toilet visits' item. 'Daily activities' was marginally preferred to 'day to day life' by patients so was changed for v13.

Version 11b	Version 12	Version 13
Item 25	Item 24	Item 24
How often did you feel that your urinary symptoms interfered with your social life?	How often did you feel that your urinary symptoms interfered with your day to day life (including social life and work outside the home)?	How often did you feel that your urinary symptoms interfered with your normal daily activities (e.g. social life, work outside the home)?
Not at all occasionally sometimes most of the time every time	Retained	Retained

Impact of nocturia and/or nocturnal voids

Patients in the additional cognitive interviews expressed a preference to be asked specifically about sleep or tiredness. This was seen as 'cutting to the chase' rather than asking indirectly if getting up at night affected day to day life. 'Daytime activities' was tested as an alternative to 'day to day life' but there was no preference expressed for either option. 'Day to day life' was retained and both items were included in version 13 as separate impact items. Their responsiveness to change will provide further information on whether both items will be retained in the final instrument.

Version 11b	Version 12	Version 13
Item 26	Item 25	Item 25
How often did you feel getting up at night to pass urine affected your day to day life?	Retained	Retained
Not at all occasionally sometimes most of the time every time	Retained	Retained

Version 11b	Version 12	Version 13
No corresponding item	No corresponding item	Item 26
		How often did you urinary symptoms prevent you from getting the amount of sleep you needed?
Not at all occasionally sometimes most of the time every time	Retained	Retained

Physical activity

Despite a floor effect of 46%, no changes were made to this item as this was felt to be a separate concept to the other impact items. Further information on its responsiveness to change will determine whether it will be retained in the final instrument.

Version 11b	Version 12	Version 13
Item 27	Item 26	Item 27
How often did you feel that your urinary symptoms affected your physical activities (e.g. walking, swimming, sport)?	Retained	Retained
Not at all occasionally sometimes most of the time every time	Retained	Retained

Way feel about self

The examples given in version 12 were agreed by most of the patients in the additional cognitive interviews as improving the clarity of the question. However, the specificity of the question was also limited to the examples given. For example, one patient highlighted concern that the examples do not capture her sense of frustration and irritation which she may associate with this question. Due to no previous specific evidence of frustration or irritation in previous CE/cognitive interviews the examples were retained in version 13.

Version 11b	Version 12	Version 13
Item 28	Item 27	Item 28
Did your urinary problem affect the way you feel about yourself?	Did your urinary symptoms affect the way you feel about yourself (e.g. embarrassment, self-confidence, self-esteem)?	Retained
Not at all occasionally sometimes most of the time every time	Retained	Retained

Embarrassment

This item was removed in version 12. The emotion 'embarrassment' was included as one of the examples in item 28, and inclusion was therefore considered duplication.

Version 11b	Version 12	Version 13
Item 29	Removed	
Did your urinary problem cause you to feel embarrassed?		
No In the morning In the afternoon In the evening At night		

Fluid intake

This item showed the lowest average bother score of the impact items although still relatively high (5.75). The item was left unchanged awaiting further responsiveness to change data.

Version 11b	Version 12	Version 13
Item 29	Item 28	Item 29
Did your urinary symptoms cause you to be careful about how much or the type of fluid you drink?	Retained	Retained
Not at all occasionally sometimes most of the time every time	Retained	Retained

Overall impact

No changes were made to this item as it was answered well, and an overall item for assessing quality of life was considered clinically useful.

Version 11b	Version 12	Version 13
Item 30	Item 29	Item 30
Overall, how much would you say your urinary symptoms interfered with your everyday life?	Retained	Retained

Appendix 8 Revisions and rationale for changes to the draft ICIQ-UAB following the interviews in the US and Japan.

Introduction

Concept elicitation interviews were conducted in the United States and Japan, detailed in chapter 7. The implications for the instrument of a number of findings were discussed by the development team in September, 2017. When making decisions surrounding the modification, inclusion or removal of items, consideration was given to the clinical utility, frequency, bother, and the spontaneity by which a concept was reported.

- Although coded as 'lower urinary tract pain' in the UK interviews, pain or discomfort, perhaps as a result of a post void residual or full bladder was a concept that was reported relatively frequently in the US and Japanese interviews. However, the location or underlying aetiology of bladder pain may be complicated. When related items were explored in the original cognitive interviews, this uncertainty in the origin or characterisation of bladder pain was the rationale for not including an item relating to this concept. However, due to the prominence of pain or discomfort in the US and Japanese findings, an item relating to this concept which was removed during the original cognitive interviews was proposed to be adapted and re-instated.
- A long urination time and impacts of hygiene and on clothing appearance in the Japanese sample were reported by one patient so were not considered sufficient to include separate items to capture these concepts.
- The increased prominence of the reporting of impact on sex-lives in the US sample could reflect cultural openness of talking about sex-related issues. However, it was recognised that side-effects of urological surgery or medication which can affect sex-drive and/or erectile dysfunction in men [217] could complicate the measurement of this concept. It was also only reported by a minority of patients. An additional impact item to measure this concept was deemed not warranted by the development team.
- Additional items or modification to the existing items in the ICIQ-UAB which measure the emotional impact (e.g. frustration, anxiety, effects on self-esteem and confidence) were considered. The development team considered that these concepts were captured sufficiently by the current impact items in the instrument.
- The financial impact of UAB reported in the US was not considered for the inclusion of an associated item, as this concept is applicable to any disease requiring healthcare, dependent on income, and would not respond to a healthcare intervention.

Following the development team meeting in September 2017, some proposed minor modifications to some items resulted in version 14. These changes were tested in further confirmatory cognitive interviews, conducted over the telephone. The following summarises the rationale for changes made to ICIQ-UAB version 13 and version 14, resulting in the final 'developmental' version 15. For most items there was no change, so details are only given for the aspects of the instrument where changes were made.

Confirmatory cognitive interview sample

Four individual confirmatory interviews were conducted to test the modifications (Table 2). The participants also participated in the additional cognitive interviews. Informed consent was taken over the phone and all were audio recorded. Participants were emailed the draft questionnaire in advance of the call. Notes were taken during the interviews using an interview schedule.

Table 2. Participant characteristics for the confirmatory cognitive interviews.

Confirmatory interview number	Additional CD interview number	CD study number	Diagnostic group	Gender	Age (years)	PVR	ISC
01	01	CD P21	DU + USI	F	70	Yes	No
02	05	CD P5	DU	M	69	Yes	Yes, every day
03	02	CD P6	DU+ BOO + DO	M	81	Yes	No
04	03	CD P4	DU + DO + BOO	M	67	Yes	No

Abbreviations: Detrusor underactivity (DU), detrusor overactivity (DO), stress urinary incontinence (SUI), bladder outlet obstruction (BOO), bladder outlet obstruction in the equivocal range (BOO-E), Post void residual (PVR), Intermittent self-catheterisation (ISC)

Final questionnaire modifications

Instructions

The development team meeting suggested a minor rephrasing of the initial introductory instructions to the questionnaire. The patients in the confirmatory interviews agreed with the changes but suggested to add 'or all of the time' in the first sentence to reflect their experience. The words 'aims to' were also removed for clarity.

Version 13	Version 14	Version 15
Many people experience urinary symptoms some of the time. This questionnaire aims to find out whether you experience symptoms associated with underactive bladder, and whether or not these symptoms have an impact on your everyday life.	Many people experience urinary symptoms some of the time. This questionnaire aims to ask you about symptoms that are associated with underactive bladder, and whether or not these symptoms have an impact on your everyday life.	Many people experience urinary symptoms, some or all of the time. This questionnaire asks you about symptoms that are associated with underactive bladder, and whether or not these symptoms have an impact on your everyday life.

Self-catheterisation

At the development meeting there was clinical value placed in capturing historical self-catheterisation. The item was therefore tested in two parts to capture this information which was accepted well by patients in the confirmatory interviews. This format was slightly changed in version 15 to include the initial 'yes/no' question as a separate item to improve clarity. However, item 3 in version 15 is to be considered for removal from the tool at a later stage, with the possible inclusion in a baseline questionnaire or case report form (CRF) as part of a study, or separately in clinical use.

Version 13	Version 14	Version 15
Item 3	Item 3	Item 3 and item 4
Over the last week, how often did you self-catheterise?	Have you ever self-catheterised? (yes/no) If so, how often did you self-catheterise over the last week?	Item 3: Have you ever self-catheterised? (yes/no) Item 4: Over the last week, how often did you self-catheterise?
Not at all less than once a day (1-6 times per week) 1-2 times a day 3-4 times a day 5 or more times a day	Retained	Retained

Daytime urinary frequency

The second response option was changed from '4-8' to '4-7' and the remaining options adjusted accordingly. This was to corroborate with clinically accepted parameters for what constitutes 'normal' number of micturitions per day. This was checked by the confirmatory interviews and was interpreted as intended.

Version 13	Version 14	Version 15	Scoring
Item 18	Item 19	Item 20	
During the day, how many times did you pass urine?	During the day, how many times did you pass urine?	Retained	
1-3 times 4-8 times 9-10 times 11-12 times 13 or more times	1-3 times 4-7 times 8-9 times 10-11 times 12 or more times	Retained	1 0 1 2 3

Incontinence

It was agreed by the development team that there was clinical utility in capturing different types of incontinence. The incontinence item in version 13 was split into two items to capture urgency, and stress urinary incontinence as separate items. These were understood and accepted well by the patients in the confirmatory interviews. One patient explained that 'physical activity, coughing or sneezing' may not be applicable to the last 24 hours. However, on discussion with the team it was accepted that although this may have the effect of exacerbating any floor effect associated with this item, it is preferable to await further evidence of the responsiveness of these items. Both items were retained.

Version 13	Version 14	Version 15
Item 8	Item 8 and item 9	Item 9 and item 10
How often did you leak urine (e.g. before you could get to the toilet, when physically active, or when you coughed or sneezed)?	Item 8: How often did you leak urine before you could get to the toilet? Item 9: How often did you leak urine when physically active, or when you coughed or sneezed?	Retained
Not at all occasionally about half the time most of the time every time	Retained	Retained

Bladder pain or discomfort

This item was adapted and re-instated as a result of the additional evidence obtained from the interviews in the US and Japan. The wording was based on previously tested versions of items in the original cognitive interviews, relating to bladder pain or discomfort. The patients in the confirmatory interviews had no difficulty answering the item. However, their interpretation could be normal discomfort when experiencing a full bladder, or bladder pain/discomfort in other circumstances. The words pain and discomfort also appeared to be interchangeable in meaning, although discomfort perhaps was perceived as being less severe. The item was retained with these caveats, with the knowledge that it is likely to be considered for removal from the tool at a later stage, following responsiveness to change data.

Version 13	Version 14	Version 15
	Item 24	Item 25
No item	How often did you feel pain or discomfort in your bladder as it filled?	Retained
	Not at all occasionally about half the time most of the time every time	Retained

Conclusion

The pilot study provided information on the measurement properties of the ICIQ-UAB. The proposed changes to the questionnaire as a result of this evidence were subjected to additional cognitive interviews which provided feedback on their facility of interpretation and understanding. Following a development team meeting and evidence from the concept elicitation interviews in US and Japan, additional proposed modifications were checked by confirmatory cognitive interviews. This resulted in a final 'developmental version' of the ICIQ-UAB (version 15). The 33-item instrument is now ready for further psychometric testing, in particular the evaluation of the sensitivity to change and derivation of a scoring system.

Appendix 9 Final developmental version of the ICIQ-UAB (v15)

Initial number **CONFIDENTIAL** DAY MONTH YEAR

Today's date

Underactive Bladder

Many people experience urinary symptoms, some or all of the time. This questionnaire asks you about symptoms that are associated with underactive bladder, and whether or not these symptoms have an impact on your everyday life.

Please write in your date of birth: DAY MONTH YEAR

Are you *(tick one)***:** Female ☐ Male ☐

1a. Have you ever been unable to pass urine <u>at all</u> and had to go to hospital to have a catheter tube inserted to drain the bladder?

no ☐ 0
once ☐ 1
twice ☐ 2
three or more times ☐ 3

b. How much does this bother you?
Please circle a number between 0 (not at all) and 10 (a great deal)

0 1 2 3 4 5 6 7 8 9 **10**
not at all a great deal

2a. Over the last 12 months, have you had a <u>urinary infection</u> for which you took medication?

no ☐ 0
unsure ☐ 1
once ☐ 2
twice ☐ 3
three or more times ☐ 4

b. How much does this bother you?
Please circle a number between 0 (not at all) and 10 (a great deal)

0 1 2 3 4 5 6 7 8 9 **10**
not at all a great deal

3. Have you ever self-catheterised?

no ☐ 0

yes ☐ 1

4a. Over the <u>last week</u>, how often did you self-catheterise?

not at all ☐ 0

less than once a day (1-6 times) ☐ 1

1-2 times a day ☐ 2

3-4 times a day ☐ 3

5 or more times a day ☐ 4

b. How much does this bother you?

Please circle a number between 0 (not at all) and 10 (a great deal)

0 1 2 3 4 5 6 7 8 9 **10**

not at all — a great deal

Please answer the following questions thinking about how your symptoms have been over the <u>LAST 24 HOURS.</u>

Over the <u>LAST 24 HOURS…</u>

5a. When ready to pass urine, was there a delay before the urine flow started?

not at all ☐ 0

occasionally ☐ 1

about half the time ☐ 2

most of the time ☐ 3

every time ☐ 4

b. How much does this bother you?

Please circle a number between 0 (not at all) and 10 (a great deal)

0 1 2 3 4 5 6 7 8 9 **10**

not at all — a great deal

Over the <u>LAST 24 HOURS…</u>

6a. When ready to pass urine, did you feel you had to concentrate in order to start passing urine?

not at all	☐	0
occasionally	☐	1
about half the time	☐	2
most of the time	☐	3
every time	☐	4

b. How much does this bother you?
Please circle a number between 0 (not at all) and 10 (a great deal)

0 1 2 3 4 5 6 7 8 9 **10**
not at all a great deal

Over the <u>LAST 24 HOURS…</u>

7a. How often were you only able to pass a small volume of urine?

not at all	☐	0
occasionally	☐	1
about half the time	☐	2
most of the time	☐	3
every time	☐	4

b. How much does this bother you?
Please circle a number between 0 (not at all) and 10 (a great deal)

0 1 2 3 4 5 6 7 8 9 **10**
not at all a great deal

Over the <u>LAST 24 HOURS…</u>

8a. How often did a few drops leak out into your underwear shortly after you had finished passing urine and had dressed yourself?

not at all	☐	0
occasionally	☐	1
about half the time	☐	2
most of the time	☐	3
every time	☐	4

b. How much does this bother you?
Please circle a number between 0 (not at all) and 10 (a great deal)

0 1 2 3 4 5 6 7 8 9 **10**
not at all a great deal

Over the <u>LAST 24 HOURS…</u>

9a. How often did you leak urine before you could get to the toilet?

not at all ☐	0
occasionally ☐	1
about half the time ☐	2
most of the time ☐	3
all of the time ☐	4

b. How much does this bother you?
Please circle a number between 0 (not at all) and 10 (a great deal)

0 1 2 3 4 5 6 7 8 9 **10**
not at all a great deal

Over the <u>LAST 24 HOURS…</u>

10a. How often did you leak urine when physically active, or when you coughed or sneezed?

not at all ☐	0
occasionally ☐	1
about half the time ☐	2
most of the time ☐	3
all of the time ☐	4

b. How much does this bother you?
Please circle a number between 0 (not at all) and 10 (a great deal)

0 1 2 3 4 5 6 7 8 9 **10**
not at all a great deal

Over the <u>LAST 24 HOURS…</u>

11a. After passing urine, how often did you have to return to the bathroom to pass urine again, within a short space of time (e.g. within 15 minutes)?

not at all ☐	0
occasionally ☐	1
about half the time ☐	2
most of the time ☐	3
every time ☐	4

b. How much does this bother you?
Please circle a number between 0 (not at all) and 10 (a great deal)

0 1 2 3 4 5 6 7 8 9 **10**
not at all a great deal

Over the <u>LAST 24 HOURS…</u>

12a. Soon after passing urine, how often did you have a <u>sensation</u> that your bladder was not completely empty?

not at all	☐	0
occasionally	☐	1
about half the time	☐	2
most of the time	☐	3
every time	☐	4

b. How much does this bother you?
Please circle a number between 0 (not at all) and 10 (a great deal)

0 1 2 3 4 5 6 7 8 9 **10**
not at all a great deal

Over the <u>LAST 24 HOURS…</u>

13a. When passing urine, how often did the flow stop and start?

not at all	☐	0
occasionally	☐	1
about half the time	☐	2
most of the time	☐	3
every time	☐	4

b. How much does this bother you?
Please circle a number between 0 (not at all) and 10 (a great deal)

0 1 2 3 4 5 6 7 8 9 **10**
not at all a great deal

Over the <u>LAST 24 HOURS…</u>

14a. How often did you strain to <u>start</u> passing urine?

not at all	☐	0
occasionally	☐	1
about half the time	☐	2
most of the time	☐	3
every time	☐	4

b. How much does this bother you?
Please circle a number between 0 (not at all) and 10 (a great deal)

0 1 2 3 4 5 6 7 8 9 **10**
not at all a great deal

Over the <u>LAST 24 HOURS…</u>

15a. How often did you strain to <u>maintain</u> your flow when passing urine?

not at all ☐	0
occasionally ☐	1
about half the time ☐	2
most of the time ☐	3
every time ☐	4

b. How much does this bother you?
Please circle a number between 0 (not at all) and 10 (a great deal)

0 1 2 3 4 5 6 7 8 9 **10**
not at all a great deal

Over the <u>LAST 24 HOURS…</u>

16a. How often did you strain towards the <u>end</u> of passing urine, to try and empty your bladder?

not at all ☐	0
occasionally ☐	1
about half the time ☐	2
most of the time ☐	3
every time ☐	4

b. How much does this bother you?
Please circle a number between 0 (not at all) and 10 (a great deal)

0 1 2 3 4 5 6 7 8 9 **10**
not at all a great deal

Over the <u>LAST 24 HOURS…</u>

17a. On average, would you say that the strength of flow of your urinary stream was…

normal (not reduced) ☐	0
a little reduced ☐	1
reduced ☐	2
very reduced ☐	3
extremely reduced ☐	4

b. How much does this bother you?
Please circle a number between 0 (not at all) and 10 (a great deal)

0 1 2 3 4 5 6 7 8 9 **10**
not at all a great deal

Over the <u>LAST 24 HOURS…</u>

18a. How often did you experience a sudden or strong need to pass urine which you were <u>unable</u> to ignore, and had to rush to the bathroom?

not at all	☐	0
occasionally	☐	1
about half the time	☐	2
most of the time	☐	3
every time	☐	4

b. How much does this bother you?
Please circle a number between 0 (not at all) and 10 (a great deal)

0 1 2 3 4 5 6 7 8 9 **10**
not at all a great deal

Over the <u>LAST 24 HOURS…</u>

19a. During the night, how many times did you have to get up to pass urine?

not at all	☐	0
once	☐	1
twice	☐	2
three times	☐	3
four or more times	☐	4

b. How much does this bother you?
Please circle a number between 0 (not at all) and 10 (a great deal)

0 1 2 3 4 5 6 7 8 9 **10**
not at all a great deal

Over the <u>LAST 24 HOURS…</u>

20a. During the day, how many times did you pass urine?

1-3 times	☐	1
4-7 times	☐	0
8-9 times	☐	1
10-11 times	☐	2
12 or more times	☐	3

b. How much does this bother you?
Please circle a number between 0 (not at all) and 10 (a great deal)

0 1 2 3 4 5 6 7 8 9 **10**
not at all a great deal

Over the <u>LAST 24 HOURS…</u>

21a. Did you find it difficult to tell when your bladder was full?

not at all	☐	0
occasionally	☐	1
about half the time	☐	2
most of the time	☐	3
all of the time	☐	4

b. How much does this bother you?
Please circle a number between 0 (not at all) and 10 (a great deal)

0 1 2 3 4 5 6 7 8 9 **10**
not at all a great deal

Over the <u>LAST 24 HOURS…</u>

22a. How often did you stay a bit longer in the bathroom after passing urine, to make sure your bladder was as empty as possible?

not at all	☐	0
occasionally	☐	1
about half the time	☐	2
most of the time	☐	3
every time	☐	4

b. How much does this bother you?
Please circle a number between 0 (not at all) and 10 (a great deal)

0 1 2 3 4 5 6 7 8 9 **10**
not at all a great deal

Over the LAST 24 HOURS…

23a. What was the longest time that you needed to spend in the bathroom trying to empty your bladder?

1-2 minutes	☐	0
3-5 minutes	☐	1
6-10 minutes	☐	2
11-15 minutes	☐	3
more than 15 minutes	☐	4

b. How much does this bother you?
Please circle a number between 0 (not at all) and 10 (a great deal)

0 1 2 3 4 5 6 7 8 9 **10**
not at all a great deal

Over the LAST 24 HOURS…

24a. How often did you go to the toilet to pass urine but were unable to urinate at all, so had to return to the bathroom to try again later?

not at all	☐	0
occasionally	☐	1
about half the time	☐	2
most of the time	☐	3

b. How much does this bother you?
Please circle a number between 0 (not at all) and 10 (a great deal)

0 1 2 3 4 5 6 7 8 9 **10**
not at all a great deal

Over the LAST 24 HOURS…

25a. How often did you feel pain or discomfort in your bladder as it filled?

not at all	☐	0
occasionally	☐	1
about half the time	☐	2
most of the time	☐	3
every time	☐	4

b. How much does this bother you?
Please circle a number between 0 (not at all) and 10 (a great deal)

0 1 2 3 4 5 6 7 8 9 **10**
not at all a great deal

Impact

The following questions relate to how these symptoms may have affected your everyday life. Please answer each question according to how often you have felt this way over the **LAST WEEK.**

Over the LAST WEEK…

26a. **How often did you make plans around the location of toilets (e.g. shopping, social outings, travelling, holidays)?**

not at all	☐	0
occasionally	☐	1
sometimes	☐	2
most of the time	☐	3
all of the time	☐	4

b. **How much does this bother you?**
Please circle a number between 0 (not at all) and 10 (a great deal)

0 1 2 3 4 5 6 7 8 9 **10**
not at all a great deal

Over the LAST WEEK…

27a. **How often did you feel that your urinary symptoms interfered with your normal daily activities (e.g. social life, work outside the home)?**

not at all	☐	0
occasionally	☐	1
sometimes	☐	2
most of the time	☐	3
all of the time	☐	4

b. **How much does this bother you?**
Please circle a number between 0 (not at all) and 10 (a great deal)

0 1 2 3 4 5 6 7 8 9 **10**
not at all a great deal

Over the <u>LAST WEEK…</u>

28a. How often did you feel getting up <u>at night</u> to pass urine affected your day to day life?

not at all	☐	0
occasionally	☐	1
sometimes	☐	2
most of the time	☐	3
all of the time	☐	4

b. How much does this bother you?
Please circle a number between 0 (not at all) and 10 (a great deal)

0 1 2 3 4 5 6 7 8 9 **10**
not at all a great deal

Over the <u>LAST WEEK…</u>

29a. How often did your urinary symptoms prevent you from getting the amount of sleep you needed?

not at all	☐	0
occasionally	☐	1
sometimes	☐	2
most of the time	☐	3
all of the time	☐	4

b. How much does this bother you?
Please circle a number between 0 (not at all) and 10 (a great deal)

0 1 2 3 4 5 6 7 8 9 **10**
not at all a great deal

Over the <u>LAST WEEK…</u>

30a. How often did you feel that your urinary symptoms affected your physical activities (e.g. walking, swimming, sport)?

not at all	☐	0
occasionally	☐	1
sometimes	☐	2
most of the time	☐	3
all of the time	☐	4

b. How much does this bother you?
Please circle a number between 0 (not at all) and 10 (a great deal)

0 1 2 3 4 5 6 7 8 9 **10**
not at all a great deal

Over the <u>LAST WEEK…</u>

31a. Did your urinary symptoms affect the way you feel about yourself (e.g. embarrassment, self-confidence, self-esteem)?

not at all	☐	0
occasionally	☐	1
sometimes	☐	2
most of the time	☐	3
all of the time	☐	4

b. How much does this bother you?
Please circle a number between 0 (not at all) and 10 (a great deal)

0 1 2 3 4 5 6 7 8 9 **10**
not at all a great deal

Over the <u>LAST WEEK…</u>

32a. Did your <u>urinary symptoms</u> cause you to be careful about how much or the type of fluid you drink?

not at all	☐	0
occasionally	☐	1
sometimes	☐	2
most of the time	☐	3
all of the time	☐	4

b. How much does this bother you?
Please circle a number between 0 (not at all) and 10 (a great deal)

0 1 2 3 4 5 6 7 8 9 **10**
not at all a great deal

Over the <u>LAST WEEK…</u>

33. Overall, how much would you say your urinary symptoms interfered with your everyday life?

Please circle a number between 0 (not at all) and 10 (a great deal)

0 1 2 3 4 5 6 7 8 9 **10**
not at all a great deal

Thank you very much for answering these questions.